

58-34 SEARCH REQUEST FORM

Attach Paper #11

Requestor's Name: Brownen Loop Serial Number: 09/715,249

Date: _____ Phone: 605-1197 Art Unit: 1636

MAILBOX: 11E12 OFFICE: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Please search the commercial nucleotide
databases for ~~sequence~~ ²⁰³⁴ 1-~~72~~ of
SEQ ID No. 1.

Please also search for ~~sequence~~ ^{nucleotides} 1-72
939-2034
and ~~sequence~~ of SEQ ID No. 1 in the
commercial nucleotide databases.

Thank you.

Edward Hart
Technical Info Specialist
STIC / Biotech
CAL 12C14 Tel: 305-9203

STAFF USE ONLY

Date completed: 11/14/02
Searcher: 11/15/02
Terminal time: _____
Elapsed time: _____
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
☒ STIC
☒ CM-1
☐ Pre-S
Type of Search
☒ N.A. Sequence
☐ A.A. Sequence
☐ Structure
☐ Bibliographic

Vendors
☐ IG
☐ STN
☐ Dialog
☐ APS
☐ Geninfo
☐ SDC
☐ DARC/Questel
☒ Other

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: January 14, 2002, 19:04:15 ; Search time 2222.52 seconds
(without adjustments)

15097,845 Million cell updates/sec

Title: US-09-715-249-1_COPY_1_2034
 Perfect score: 2034

Sequence: 1 atgcgacccctccggyacgac.....acatcgttcggaagcgacg 2034

Scoring table:

	IDENTITY_NUC	
Gap00 10 0		Gap00 1 0
Gap00 10 0		Gap00 1 0

Searched: 1472140 seqs, 8248589755 residues

```
.l number of hits satisfying chosen parameters: 2944280
```

Maximum DB seq Length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1:	gb_hbg.*
2:	gb_in.*
3:	gb_cm.*
4:	gb_ov.*
5:	gb_ph.*
6:	gb_pl.*
7:	gb_pr.*
8:	gb_ro.*
9:	gb_sts.*
10:	gb_sy.*
11:	gb_vl.*
12:	em_fun.*
13:	em_hum.*
14:	em_in.*
15:	em_com.*
16:	em_ov.*
17:	em_ph.*
18:	em_pl.*
19:	em_ro.*
20:	em_sts.*
21:	em_sy.*
22:	em_vl.*
23:	em_hbg.*
24:	em_in.*
25:	em_cm.*
26:	em_ov.*
27:	em_ph.*
28:	em_pl.*
29:	em_ro.*
30:	em_sts.*
31:	em_sy.*
32:	em_vl.*
33:	em_hbg.*
34:	em_in.*
35:	em_cm.*
36:	em_ov.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2034	100.0	3633	6	AX179384	AX179384 Sequence
2	2034	100.0	5532	6	HS066089	HS066089 Sequence
3	2034	100.0	5532	9	HSECFPRE	X00588 Human mRNA
4	1918	94.3	2643	9	HUMCFPRS	K03193 Human abert
5	1876.8	92.3	2864	9	AF125253	AF125253 Homo sapi
6	1606.4	79.0	2400	9	HSECF205	X00663 Human mRNA
7	1484.2	73.0	4188	10	HMBCGR	X78987 M.musculus
8	1484.2	73.0	5935	10	AF275367	AF275367 Mus muscu
9	1484	73.0	4194	10	RATBEGGR	X57394 Rattus norv
10	1481	72.8	2301	10	HMBCGR	M59698 M.musculus
11	1481	72.8	3704	10	MM003425	M03425 Mus musculu
12	1416.8	67.7	2618	10	AF124513	AF124513 Mus muscu
13	1207.8	55.4	1593	6	HS048722	U48722 Human epide
14	1207.8	55.4	1866	6	180039	180039 Sequence 1
15	1207.8	59.4	1868	9	HS095089	U95089 Human trinc
16	1013.2	49.8	2400	5	CHRGKGR	M20386 Chicken epi
17	992	48.8	1958	10	AF187818	AF187818 Rattus no
18	764.2	37.6	2243	5	CHRGKGR	M77637 Gallus gall
19	563.2	27.7	5766	5	X0053471	M76371 Xiphophorus
20	552.4	27.2	4138	5	XSMTK	X18691 Xiphophorus
21	450.4	22.1	4060	10	AF041838	AF041838 Rattus no
22	431	21.2	5484	9	HUMREYKIN	L07668 Homo sapien
23	431	21.2	5501	6	AR040740	AR040740 Sequence
24	431	21.2	5555	6	AR040741	AR040741 Sequence
25	419	20.6	3768	6	AX060704	AX060704 Sequence
26	419	20.6	4530	6	121124	121124 Sequence 9
27	419	20.6	4530	6	159745	159745 Sequence 9
28	419	20.6	4530	6	HUMHERA	M11730 Human Lyrus
29	419	20.6	9274	6	AX060703	AX060703 Sequence
30	417.4	20.5	3768	6	AR034479	AR034479 Sequence
31	417.4	20.5	4473	6	AR080259	AR080259 Sequence
32	417.4	20.5	4473	6	HSEB2R	X03363 Human c-erb
33	410.6	20.2	4424	10	RNU29339	U29339 Rattus norv
34	409.4	20.1	4879	9	HUMGCFRBB3	M29366 Human epide
35	409.4	20.1	4879	11	G28554	G28554 human SRS S
36	409.4	20.1	4905	6	AR047857	AR047857 Sequence
37	409.4	20.1	4905	6	116885	116885 Sequence 3
38	409.4	20.1	4975	6	HUMHERA	M34309 Human epide
39	405	19.9	4062	10	HANNEU	D16295 Syrian goldf
40	402.2	19.8	3955	6	121129	121129 Sequence 14
41	402.2	19.8	3955	6	159750	159750 Sequence 14
42	402.2	19.8	3955	10	RNEEUR	X03362 Rat mRNA id
43	400.6	19.7	2385	6	AR082744	AR082744 Sequence
44	400.6	19.7	2385	6	AR093963	AR093963 Sequence
45	400.6	19.7	2385	6	AR143949	AR143949 Sequence

OY	1	atgcgacccctccgggaagcgcggggaagcgctccctgggcctctctgctcgccctctgcgcg	60
Db	1	atgcgacccctccgggaagcgcggggaagcgctccctgggcctctctgctcgccctctgcgcg	60
OY	61	gcgagctcgggctctcgaggaagaaagattctgcgaagcgacgagtaaaagctcacgcg	120
Db	61	gcgagctcgggctctcgaggaagaaagattctgcgaagcgacgagtaaaagctcacgcg	120
OY	121	ctggagacatttggaaatcattcttcctaacgctccagagagatctaaataactgtagggt	180
Db	121	ctggagacatttggaaatcattcttcctaacgctccagagagatctaaataactgtagggt	180
OY	181	gtcccttgggaatttggaaattacctaactgtagcaaggaattatgaactcttcctcctaag	240
Db	181	gtcccttgggaatttggaaattacctaactgtagcaaggaattatgaactcttcctcctaag	240
OY	241	accatccagaggatgcgcgttatgtccctaattgcctccaaagcgtagggagaaattcc	300
Db	241	accatccagaggatgcgcgttatgtccctaattgcctccaaagcgtagggagaaattcc	300
OY	301	ctggaaaacctcgacatcatcagaggaatatagttaactagaataattcctatgccttagca	360
Db	301	ctggaaaacctcgacatcatcagaggaatatagttaactagaataattcctatgccttagca	360
OY	361	gtctctctcaaatatgatgatcaataaaccggagcttgaaggagcttgcctcatgagaattta	420
Db	361	gtctctctcaaatatgatgatcaataaaccggagcttgaaggagcttgcctcatgagaattta	420
OY	421	caggaaatctctgacatgtagccgctgtaggtcttcagaaacaaacctgacctgtgtcaacgttgag	480
Db	421	caggaaatctctgacatgtagccgctgtaggtcttcagaaacaaacctgacctgtgtcaacgttgag	480
OY	481	agatattcagtgtagcggaagatactgtaacagatgactcttcagagcaaatgtcgaatgttc	540
Db	481	agatattcagtgtagcggaagatactgtaacagatgactcttcagagcaaatgtcgaatgttc	540
OY	541	cagaacaaaccttggagagctctgcgaagaagtgtgataccaagctgtcccaattgagagctgtag	600
Db	541	cagaacaaaccttggagagctctgcgaagaagtgtgataccaagctgtcccaattgagagctgtag	600
OY	601	gtgtcagagagaggaactctgcagaatactgaccaaatacattgtgtccagcagtgctcc	660
Db	601	gtgtcagagagaggaactctgcagaatactgaccaaatacattgtgtccagcagtgctcc	660
OY	661	ggagcgttcgcgttggaaagtctccccaagtgaactctgcacaaacaaagctgtgtctgacagctgc	720
Db	661	ggagcgttcgcgttggaaagtctccccaagtgaactctgcacaaacaaagctgtgtctgacagctgc	720
OY	721	acaagcccccgagagagagactgactgtgtctgtgcgcaaatctccgagacgaagccacgtgc	780
Db	721	acaagcccccgagagagagactgactgtgtctgtgcgcaaatctccgagacgaagccacgtgc	780
OY	781	aaggaacactgcgccccaactcaatgactctcaaaccccaacagctgcagatagtagtgtagac	840
Db	781	aaggaacactgcgccccaactcaatgactctcaaaccccaacagctgcagatagtagtgtagac	840
OY	841	cccgagggcaaatcaagactcttggtagccactgagctgtagaagatgtctcccgtaattatgtg	900
Db	841	cccgagggcaaatcaagactcttggtagccactgagctgtagaagatgtctcccgtaattatgtg	900
OY	901	gtgaaagatacagagcttgcgtccgaagcttggtagcggaagcttggtagatggagga	960
Db	901	gtgaaagatacagagcttgcgtccgaagcttggtagcggaagcttggtagatggagga	960
OY	961	gaacgagctccgcaagctgttaagaagtgcgaagagccttgcgcgaagatgtatcagaata	1020
Db	961	gaacgagctccgcaagctgttaagaagtgcgaagagccttgcgcgaagatgtatcagaata	1020

Dp	961	GACGGCGTCCGAAAGTGTGAAGAAAGTGGAAAGGGGCTTGGCGCAAAAGCTGTAAACGGAATA	1029
Qy	1021	ggtatgtgtgaaatcttaaaagctcaatcttccataaatgtctaagaaataataaacattcaaa	1080
Dp	1021	GGTATTGGTGAATTTAAAGACTCACTCTCCATTAAGTACGAATAATTAAACACTTCCAA	1080
Qy	1081	aactgcacctccatacagtgagatctctccaaatcttcagtgagcttcgaaggtgaaccc	1140
Dp	1081	AACTCACACTCCATCAAGTGGGAGATCTCCACATCTCTCCGTGGCATTTAAGGGGTACTCC	1140
Qy	1141	ttccacacatactctctctcttggaatccacaggaacttgatattcttaaaacgttaagaa	1200
Dp	1141	TTCAACATTACTCTCTCTCTCTGATCCACAGAACTGGATTTCTAATAAACGTTAAGTAA	1200
Qy	1201	atcaacaggytcttctgcatctcaagctctgagctctgaaacaaagaaacggaacctccatgcctt	1260
Dp	1201	ATCAAGGGTATTGTGATTCAGATTCAAGGCTGGCCCTGAAAMCAGAGACGACTCCATGCTTT	1260
Qy	1261	gagaacccatagaatacatagccggcagagaaacaaacaaatgagctcagttcttccttgagct	1320
Dp	1261	GAGAAACCTAGAAATCATACCGCGGAGACCAACAAATGCTCATTTTCTCTTGCACTC	1320
Qy	1321	gtcagctctgaacataacatcccttgaggatcagctccctcaagagaaataagtatgagat	1380
Dp	1321	GTCACCCCTGAACATTAACATCTTTGGATTACGCTCCCTCAAGGAATTAAGTATGGAGAT	1380
Qy	1381	gtgataaelttcaagaaacaaataatctgtcgalgcaatatataacttgaaataaactg	1440
Dp	1381	GTGATAAATTTACGAAACAAATAATTTGTGCTATGCAATATCAATTAACCTGGAAAACTG	1440
Qy	1441	tttggagccctccggttcagaaacaaatctataagaaagaaggtgaaataacgctctcaag	1500
Dp	1441	TTTGGAGCCTCCGGCTCAGAAACCAAAATTAAGCAACAGAGGTGAAACACGCTGAAAG	1500
Qy	1501	gccaacagcccaagctctgcaatgctctgtcccccagagagctctctggagcccgagagccc	1560
Dp	1501	GCCACAGGCGACGTCCTCCATATGCCCTTGTGCTCCCCGAGGCTGCTGGGCGCCGAGACC	1560
Qy	1561	agggaactgcgctctctcttgcggaatgtgcaagccgagcagaggaatctgtgagaagaatgcaag	1620
Dp	1561	AGGGAACGTGCTCTCTCTCTCCGGAATGTACGCCGAGAGGAATGCTGACAAAGTGCAAAG	1620
Qy	1621	ctctctggaaggttgagccaaaggaagctctgtgagaactctcgaatgcatatacagtgccaccca	1680
Dp	1621	CTTCTGGAGGGGTGAGCCAAAGGAGTTTGTGGAGAACTGTGAATGCATACACTGCCACCCA	1680
Qy	1681	gagtgccctgcctcaagggccatgaaatcacctctcaacaggaacggagacaggaataactgctatc	1740
Dp	1681	GAGTGCTGCTCAAGGCCATGAACATCACTCTGCACAGGACGGGACCGAGCAACACTGTATC	1740
Qy	1741	caagtgtgcccaatacattgaacggtcccccactgtgcgtcaagaactctgcccgaagagatg	1800
Dp	1741	CAGTGCTCCCACTAATTTAGATTCAGGCCCCCACTCGTCAAGACTGTGCCCGGACGAGTCAATG	1800
Qy	1801	ggaataaacaacaaacccctggtctctggaagtaacgaagccgcggcgcaatgtgtgcaactgtgc	1860
Dp	1801	GGAAAAAACAACACCCCTGCTGTGAAGTACGCCAGACGCCGGCCATGTGTGCCACTCTGTC	1860
Qy	1861	catccaactgtcacactaaggaatgcacttggccaggltcttgaagagctgtccaaagaaatggg	1920
Dp	1861	CATCCAAACATGCACTACAGGATCCACTGTGGGCCAGTCTTTGAAGAGCTGTCCAAACAAATGGG	1920
Qy	1921	cctaagaatcccgatccatctgcacacttggaatgagtgtggagggccctccactctgtgcactgtgtg	1980
Dp	1921	CCTTAAGATCCCGTTCAMTGGCCACTGGGATGTGTGGGGGCCCTCTCTTGTGCTGTGTGCTG	1980
Qy	1981	gcccctggtggtatcggtcctctctcaatgcaagggcgccacatcglttgggaagcgcaag	2034
Dp	1981	GCCCTGGGGATCGGCTCTTCAATGCAAGCGCCACATCTTTGGGAAGCGCACG	2034
RESULT 2			
AR06089			

Query Match	Best Local Similarity	100.0%	Score 2034	DB 9	Length 5532
Matches 2034	Conservative	0	Mismatches	0	Indels
Gaps	0				
0Y	1	atgcgacccctcgcggaagccgcgggacgcgcctctgctgcctcgtcgtcgtcgtcgtccgcg	60		
Db	187	ATGCGACCCCTCCGCGAGCGCGCGGAGCGCTCCGCGCTGCGCTCGCGCTCCGCG	246		
0Y	61	gcgcgctgcggctctgcgaagaaagaagcttgcgaagcagcaatlaacaagctcacgcgac	120		
Db	247	GCGAGTGGGCTCTGAGAGAAAAGAAAGTTTGCCAAAGCAGCAATTAACAAGCTCACGCA	306		
0Y	121	ttggagcatttgaagatcaattctctcagcctcagagagatgltcaataacvtgtgagtg	180		
Db	307	TTGGGCACTTTTGAAATCATTTTTCACCCCTCCAGAGAGATCTTCAATGTGAGGTG	366		
0Y	181	gtccctgcgaatttgcgaatlaaccatactgacgcgagagatlatgtacttctctctaaag	240		
Db	367	GTCCTTGGAATTTGGAAATTTACATATGTCGAGAGAAATTTATGATCTTCTTTAAAG	426		
0Y	241	accatccagagatgagcvtgttatagtctctcatctgcctcaacagatggagcgaattcct	300		
Db	427	ACCATCCAGAGGTGAGCTGGTTATGTCTTCATTTGCCCTTCAACACAGTGGCAATTCCT	486		
0Y	301	tttgaataaacctgcagaatcaatcagagaaatgtaactcagaaatctcctagcttagca	360		
Db	487	TTTGAATAACCTGCAGATCATCAGAGAAATATGTACTACAGAAATTCCTATGCTTAGCA	546		
0Y	361	gtcttataactataatgcgaataaacaacgcgagctgaagagagctgcccataagaaatla	420		

```

|||||
Db 547 GTCTATCTAATATGATGCAATTAATAACCGACTGAGAGAGCTGCCATGAGAAATTTA 606
Oy 421 cagaaatctctatgagcggtgaggttcagaaacccctgcccgttcgaagtgag 480
Db 607 CAGGAATCTCGATGCGCGCGCTGCGCTTCCAGAACACCCCTGCTGTGCAACGTGAG 606
Oy 481 agcatcagtcgaggacatlagtcagcagtgacattctcgaacacatgctgagcttc 540
Db 667 AGCATTCACAGTGGCGGACATAGTCACACACTTTCACACACATGTCATGAGACTTC 726
Oy 541 cagaaacacttgagcagctgagcgaagtgatgacacgctgcccgaatggagctgtcg 600
Db 727 CAGAACCACTGGGCAAGCTCCAAAAGTGTGATCCAACTGTGCCCAATGGAGAGCTGCTG 786
Oy 601 ggtgcaagagagagaaactgagcaaaactgacaaatcatctgtgcccagcagtgcttc 600
Db 787 GGTGACAGAGAGAGGAACTGCGCAGAAATGAAATCATCTGTGCGCCACAGCTGCTCC 846
Oy 661 gggcgtcagctgagcaagtcgcccaagtgacctgctcgaacaaagtgctgagctgctc 720
Db 847 GGGCGCTGCGGTGCAAGTCCCGCAGTGAAGCTGCGCACAAACCAAGTGTGCTGAGGCTGC 906
Oy 721 acaagccccgggagagagcagctgctgtctgcccgaattccgagagagagcagctgctc 780
Db 907 ACAGGCCCCCGGAGAGCGACTGCTGTGCTCCGAAATTCGAGAGAGAGAGAGAGAGAG 966
Oy 781 aagagacacgtgccccacacacacacacacacacacacacacacacacacacacacac 840
Db 967 AAGGACACTGCCCCCAGTATGCTCTACAAACCCACACACACACACACACACACACACAC 1026
Oy 841 cccgaagagagaaatcagcttggtgacacacacacacacacacacacacacacacacac 900
Db 1027 CCGGAGGCAAAATACACTTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1086
Oy 901 gtgacagacacacacacacacacacacacacacacacacacacacacacacacacacac 960
Db 1087 GTGACAGATCACGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1146
Oy 961 gacgagctgagcaagtgtaagagtgagagagccttgccgaagagtgctgaacagata 1020
Db 1147 GACGCGCTGCGCAAGTGTAAAGAGTGCAGAAAGGCTGCGCAAGTGTGTAAGAGATA 1206
Oy 1021 ggtatggtgagatttaagacacacacacacacacacacacacacacacacacacacac 1080
Db 1207 GGTATGTTGTAATTAAGACTCACTCTCATTAATGATGCAATTAATTAATTAATTAAT 1266
Oy 1081 aactgacacacacacacacacacacacacacacacacacacacacacacacacacacac 1140
Db 1267 AACTGACACTCTCATGACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1326
Oy 1141 ttcacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1200
Db 1327 TTGACACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1366
Oy 1201 atcacaaggttcttgcagatcagagcttgccgcaaaacagagacacacacacacacac 1260
Db 1387 ATCACAGAGTTTGTGATTCAGGCTTGCGCTGCAAAACAGAGAGAGAGAGAGAGAGAG 1446
Oy 1261 gagaacacacacacacacacacacacacacacacacacacacacacacacacacacac 1320
Db 1447 GAGAACTGAGAAATCATACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
Oy 1321 gtcagcctgagacataacatccttgagattagctccctcaagagagataagatagagat 1380
Db 1507 GTACACCTGAGACATACATCTCTGAGATTCAGCTGCTCTCAAGAGAGATTAAGTATGAGAT 1566
Oy 1381 gtcacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1440
Db 1567 GTGATTAATTCAGAAACAAATTTTGTGCTATGCAAAATTAACATTAACGAGAGAGAGAG 1626
Oy 1441 ttgagagacacacacacacacacacacacacacacacacacacacacacacacacacac 1500

```

```

Db 1627 TTGCGACCTCCGCTGCAAGAAACCAAAATTAAGCAACAGAGTGAAAAACACTGCAAG 1686
Oy 1501 gccacagagcaggtctgacatgctgctgctgctgctgctgctgctgctgctgctgctgct 1560
Db 1687 GCCACAGCGCAGGTCTGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1746
Oy 1561 aggaagctgctctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1620
Db 1747 AGGACATGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1806
Oy 1621 ctctgagaggtgagcagagagagagagagagagagagagagagagagagagagagagagag 1680
Db 1807 CTCTGAGAGCTGAGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1866
Oy 1681 gagtgcctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1740
Db 1867 GAGTGCCTGCTGAGGCAATGAGCAGTACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1926
Oy 1741 cagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1800
Db 1927 CAGTGTGCTGCTGAGGCAATGAGCAGTACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986
Oy 1801 ggaagaaacacacacacacacacacacacacacacacacacacacacacacacacacacac 1860
Db 1967 GAGAAACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 2046
Oy 1861 catcacaacacacacacacacacacacacacacacacacacacacacacacacacacacac 1920
Db 2047 CATCAACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 2106
Oy 1921 cctaagatccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1980
Db 2107 CCTAAGATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2166
Oy 1961 gccctgagagagagagagagagagagagagagagagagagagagagagagagagagagag 2034
Db 2167 GCCCTGGAGATGCGCTCTTCAATGAGAGGCGCACATGCTTGGAGAGGAGAG 2220

RESULT 4
HUMEGFRS 2643 bp mRNA PRI 07-NOV-1994
LOCUS Human aberrant (short) epidermal growth factor receptor mRNA,
DEFINITION complete cds.
ACCESSION K03193.1 GI:181984
VERSION K03193.1
KEYWORDS epidermal growth factor; epidermal growth factor receptor; erbB
oncogene;
SOURCE Human epidermoid carcinoma cell line A431, CDNA to mRNA, clone
PE15.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2643)
AUTHORS Merlino, G.T., Ishii, S., Whang-Peng, J., Knutsen, T., Xu, Y.-H.,
Clark, A.J., Stratton, R.D., Wilson, R.K., Ma, D.-P., Roe, B.A.,
Hunts, J.H., Shimizu, N., and Pastan, I.
TITILE Structure and localization of genes encoding aberrant and normal
epidermal growth factor receptor RNAs from A431 human carcinoma
cells.
JOURNAL Mol. Cell. Biol. 5 (7), 1722-1734 (1985) probably also shared
MEDLINE 85267689
COMMENT A draft entry and printed copy of this sequence were kindly
provided by G. Merlino (07-OCT-1985).
FEATURES
source
1. 2643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7p13-p12"
<1. >2643
/feature="EGFR"
/note="EGFR(S) mRNA: 600-120-610"
1. 2643
gene

```


[illegible]

Db	1986	CAGTGTGCCACTACATGTAATGGAASGGCCGCCACCTGCCTGCAAAAGACTGTGCCCGGACGAGTGATG	2045
Oy	1801	ggagaaaacaaaccctgcttgcgaagtgcagaagcgcggcgacatgttgccacctgtgc	1860
Db	2046	GGAATAAACACACCCTGTGTGTGGAAAGTACGACAGCAGCGCCGACCATGTGTGCCACTGTGC	2105
Oy	1861	cattccaactgcacctacgcy	1880
Db	2106	CATCCAACCTGCACCTACCG	2125
RESULT	6		
HSEGF01	LOCUS	HSEGF01	2400 bp mRNA PRI 20-MAR-1991
DEFINITION		human mRNA fragment for epidermal growth factor (EGF) receptor.	
ACCESSION		X00663	
VERSION		X00663.1	GI:31112
KEYWORDS		cellular oncogene; epidermal growth factor receptor; erbB cellular oncogene; oncogene.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
AUTHORS		1 (bases 1 to 2393) Xu,Y.H., Ishii,S., Clark,A.T., Sullivan,M., Wilson,R.K., Ma,D.P., Roe,B.A., Mellino,G.T. and Pastan,I.	
TITLE		Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs overproduced in A431 carcinoma cells	
JOURNAL		Nature 309 (1997), 806-810 (1984)	
MEDLINE		84245835	
REFERENCE		2 (bases 1 to 2400)	
AUTHORS		Merlino,G.	
TITLE		Direct Submision	
JOURNAL		Submitted (28-MAY-1985) to the EMBL/GenBank/DBJ databases	
COMMENT		Data kindly reviewed (28-05-1985) by Merlino G.	
FEATURES		Location/Qualifiers	
source		1..2400	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		<1..>2400	
		/note="EGR (1 is 2nd base in codon)"	
		/codon_start=1	
		/protein_id="CAA25282.1"	
		/db_xref="GI:4378981"	
		/db_xref="SWISS-PROT:P00533"	
		/translation="ILRHCAVFPSPNPALCNVESIQMDIYSSPFLSNMSMPDNHLGS COGDPSGNSGSCAGENGCKLTKITLQAQCSRRGSGSPDCHNCOAAGCGPR ESGVICRKRREAFCTKTCGPGLTLNPTTYNMTPRGYSFGACCYKKCRARYVVTV DHSCVRACGADSYTEMEDGYRKCKCBEPCKRYNGTIGERDSLSTINATIKHKR NCTSIQSDILLPLVFGRDSFTHPPLDPDELIDITKTVELTILQMPENRNTDLH A FENLEILRGTRKHQHGSLAVVLSLTSLGISRLKEISQDVLI ISGNNLCTANTYIN WKLPFSIQOKTK IISNKGENSKATGVQAHMALSPEBCWPPRCVASCRAWSGREE CVKRNLLGGEPRFEVENSBC IQCHEPCIPAMNMTCTGGPNCTOC AHY IDGPHVC CTPDAGMEENNLTVMKYADAGVCHLCHNCTYCCTGPLEGCPINGPKIESLATGM VTGALLLVVALGIGLFMRBHRIYRKRTLRLIDRELVEPLRPSEGAPNALIRLIK ETFERKIKVLGSGAGTGYVGLMIPGEVKVLPVALIKELREATSPANKETLDENVYVA ASYDNHYCRGLGCTISTVOYLITOLMPGCLIVREHNDISQYLLNNCYOIAAG MYNIIEPRLYHRDI LAARNVLYKTPQHVKTDPGLAKILGAEKEYNAEGVPYTNMA LESILRHITTHOSDWMSIGVYVELMTFFSRKFDIIFASEISSLKERLPPQPICT ID"	
BASE COUNT		615 a 655 c 659 g 471 t	
ORIGIN			
Query Match		79.0%:	Score 1606.4; DB 9; Length 2400;
Best local Similarity		99.9%:	Pred. No. 0;
Matches 1607; Conservative		0; Mismatches	1; Indels 0; Gaps 0;
Oy	427	atccgcgatgycgectgcygttcagcaacaaccctgcctctgtcacacytgagagacate	486
Db	1	ATCCTTCATGCGCGCCGTGCGCTTCAACAACAACCTTCCTCTGCAAGCTGGAAGACATC	60
Oy	487	catgtgcggagacatagtcacgaatgacttttcagcaaacatgtcgaibytgatatccatccgaac	546


```
/db_xref="Swiss-Prot:001279"
/translation="MRPS27A1TLLVLTALCAAGALEKKVCOGTSNRLTOLGTF
EDHLSLQRMNCEVYLGILTYVORNYDLSEFKTIOEVAGVLLALVTERPLE
NQIILRGALYENTYALALISNGTNRGLRLEPARNQELILEAVFSNPLILCMD
T10WMDIVONFNSNMSMDIOSHPSCPCDPSCMGWGENOKRLTKITCAQO
CSHRCGRSPDCSHNOCAAGCTCPRESCLICQKFODEATCKDTCPMLNPTPYO
MDVNECKYSFGATCVKCRNRYVTDHSGSCVAGSGLPYVEVEDGIRKCKDPCR
KYVNGIGIGEFKDTLSINATNKHKKYCALISGLHLLPAPKDFSTRPLDPEL
ELIKTYKRTGILLIOMDPNMTLHAPNLEITFGPKTQKQSCFSAVYVITSLG
RSKTEISDGVYIISGRNLCYANTINKKLETRPQKTIINNAREDCANVHNGN
LCSBECMGEPFROCVSNVSRGECVKEKILIEGEEVEVSECTQCHPELQOA
MNTTGTGDPNCTIQAHITDPCVCTKCPACIMENNTLWKTAKMNRHILKRLR
TYCAGPCLQCEVWSPKIPSLIATGVGLLEIVVALDGLFPMNRHILKRLR
RLDERLEVERLTPSGEPANQHLRLKTEPERKIVGSAFETVGLMIPDEGKY
KIPVAKELREATSPKANKELIDEAVANASVDNVCALGICLSTVOILTOLEMG
CLDYVERHKNDISQYILMWCYQJAKGNVLEDRILVHRLAARNVLYTPQHVLT
DEGLAKLGAEEKEEYHAEQKVPILKMALESLIRITTHOSDMSYGVYWEMLTFS
KPYDGIIPASDISILEKBERLPOPICTIDYVMTIVKCMWIMDADSRPKRELIERS
MAROPRIVIOGDERMHLSPPTDSNFRALMEDEMDVDYDADRYLROQEPNPS
TSRPTLSLSATSNSTYACINRNGSCVYKEDATLORYSPDPGATEDNIDPATL
VPEVYVOSVPRRAGSYONPYHNOPLIPAFORDLHTONRPNNAKREPILTAOPTC
LSSGRNSPALMIQKSHONSMDPDIQDFPKEKTRNGIFKGPJAEVMEYLVAPPS
SELFIA"
60.131
sig_peptide
conflict
/citation="(2)
/replace="C"
115.119
/citation="(3)
/replace="ca"
425
conflict
/citation="(3)
/replace="B"
1784
/citation="(2)
/replace="g"
1067 c 1035 g 930 c
BASE COUNT 1156 a 1067 c 1035 g 930 c
ORIGIN
Query Match 73.0% Score 1484.2 DB 10: Length 4188;
Best Local Similarity 83.4% Pred. No. 0;
Matches 1700; Conservative 0; Mismatches 333; Indels 6; Gaps 1:
QY 1 a b c d e f g h i j k l m n o p q r s t u v w x y z
Db 60 A T C G A C C C T C A G G A C C G C A G A A C C A C A C T G C T G T G T C T G A C C G C C T C T G C G C
61 g c g a g t c g g c t c g g a g a a g a a g t c t g c a a g a c g a g t a g a a g c t a c g c a g 120
120 G C A G G T G G G C C T T G G A G G A A A A G A A G T C G C C A A G C A A G T A A C A G C C T C A C C C A A 179
QY 121 t t g a g a c t t t g a a g a l c a t t t c t c a g c t c g a g a g t t t c a a t a c t t g a g a g 180
Db 180 C T G G C A C T T T T G A G A C C A C T T T C T G A C C T G C A G A T G T A C A C A C A C T G T A A G T G 239
QY 181 g t c c t t g g a a t t t g a a a t t a c t a t g c a g a g a a t a t a t a t t c t t c t t a a g 240
Db 240 G T C C T T G G A A C T T G A A A T T A C T A T G T C A A A G A A T T A C A C C T T T C T T T A A G 299
QY 241 a c a t c c a g a g a g t g t g t a t a t c t c a t t c c t c a a c a c a g t g a g a a t t c c t 300
Db 300 A C C A T C C A G A G A G T G G C G C G C T A T G T C C T C A A C C G T A A C C G T A A G A A G A A T C C T 359
QY 301 t t g a a a a c t t c a g a t c a g a g a a t a t a t a t a g a a a a t t a c t a t a g c t t a g a 360
Db 360 T T G G A A A C A C T C A G A T C A T C A G G G A A A T G C T T T A T A A A A C A C C T A T G C C T T A G C C 419
QY 361 g t c t a t a c t a t a t a t a a a c c g a c t g a a g a c t g c c a t a t a g a a a t t a 420
Db 420 A T C C T G T G A A C T A T G A A A C A A C A A C G G C T T A G G A A C T C C C A T C G G A A C T T A 479
QY 421 c a g a a a t c t c g a t g c g c t g c g t t c a g a a a c a c c t g c c t t g c a a g t g a g 480
/d_b_xref="Swiss-Prot:001279"
/translation="MRPS27A1TLLVLTALCAAGALEKKVCOGTSNRLTOLGTF
EDHLSLQRMNCEVYLGILTYVORNYDLSEFKTIOEVAGVLLALVTERPLE
NQIILRGALYENTYALALISNGTNRGLRLEPARNQELILEAVFSNPLILCMD
T10WMDIVONFNSNMSMDIOSHPSCPCDPSCMGWGENOKRLTKITCAQO
CSHRCGRSPDCSHNOCAAGCTCPRESCLICQKFODEATCKDTCPMLNPTPYO
MDVNECKYSFGATCVKCRNRYVTDHSGSCVAGSGLPYVEVEDGIRKCKDPCR
KYVNGIGIGEFKDTLSINATNKHKKYCALISGLHLLPAPKDFSTRPLDPEL
ELIKTYKRTGILLIOMDPNMTLHAPNLEITFGPKTQKQSCFSAVYVITSLG
RSKTEISDGVYIISGRNLCYANTINKKLETRPQKTIINNAREDCANVHNGN
LCSBECMGEPFROCVSNVSRGECVKEKILIEGEEVEVSECTQCHPELQOA
MNTTGTGDPNCTIQAHITDPCVCTKCPACIMENNTLWKTAKMNRHILKRLR
TYCAGPCLQCEVWSPKIPSLIATGVGLLEIVVALDGLFPMNRHILKRLR
RLDERLEVERLTPSGEPANQHLRLKTEPERKIVGSAFETVGLMIPDEGKY
KIPVAKELREATSPKANKELIDEAVANASVDNVCALGICLSTVOILTOLEMG
CLDYVERHKNDISQYILMWCYQJAKGNVLEDRILVHRLAARNVLYTPQHVLT
DEGLAKLGAEEKEEYHAEQKVPILKMALESLIRITTHOSDMSYGVYWEMLTFS
KPYDGIIPASDISILEKBERLPOPICTIDYVMTIVKCMWIMDADSRPKRELIERS
MAROPRIVIOGDERMHLSPPTDSNFRALMEDEMDVDYDADRYLROQEPNPS
TSRPTLSLSATSNSTYACINRNGSCVYKEDATLORYSPDPGATEDNIDPATL
VPEVYVOSVPRRAGSYONPYHNOPLIPAFORDLHTONRPNNAKREPILTAOPTC
LSSGRNSPALMIQKSHONSMDPDIQDFPKEKTRNGIFKGPJAEVMEYLVAPPS
SELFIA"
60.131
sig_peptide
conflict
/citation="(2)
/replace="C"
115.119
/citation="(3)
/replace="ca"
425
conflict
/citation="(3)
/replace="B"
1784
/citation="(2)
/replace="g"
1067 c 1035 g 930 c
BASE COUNT 1156 a 1067 c 1035 g 930 c
ORIGIN
Query Match 73.0% Score 1484.2 DB 10: Length 4188;
Best Local Similarity 83.4% Pred. No. 0;
Matches 1700; Conservative 0; Mismatches 333; Indels 6; Gaps 1:
QY 1 a b c d e f g h i j k l m n o p q r s t u v w x y z
Db 480 C A G A A A T C C T A A T T G G T G C T G T C C A G A T T C C A C A C C C A C C A T C T C T C A A T A T G A T 539
QY 481 a g a t t c c a g t g c g g a c a t a g t c a g c a g t a c t t c t c a g a c a c a t a t c a t a g a c t t c 540
Db 540 A C A T C C A C T G A G G A C A T G T C T C A A A C G C T T T A T A G A C A A C A T G T C A A T G A C A T T A 599
QY 541 c a g a a c c a c t t g g a g t t g c a a a g t t g t c a a g c t t c c a a t t g a g a c t t g 600
Db 600 C A A G C C A T C C A G A G A G T T G C C C A A A T G A T C C A A G T G C C A A T G A A G A C T G C T G 659
QY 601 g t g c a g a g a g a g a a c t t g c a g a a c t g a c a a a t a t c t t g c c a c a g c a g t g c c 660
Db 660 G G A G G A G A G A G A G A A C T G C A G A A A T T G A C C A A A T C A T C T G C C C A G A C A T G T T C C 719
QY 661 g g g c g t g c g t g c a a g t c c c c a g t a c t g c t c a a a c c a g t g t g c t g a g g c t g c 720
Db 720 C A T C G T G T C G T G C G A G T C C C C A G T G A C T G T G C A A A G T T C A A G A T G A G C C A C A T C 779
QY 721 a c a g g c c c c g g a g a g a c t g c t g t c t c g c a a t t c c g a a c a g c c a g t g c 780
Db 780 A C G G G C C C G A G A G T A C T G T G T G T G C A A A G T T C A A G A T G A G C C A C A T C 839
QY 781 a a g a c a c t g c c c c a c t a t a t g c t c t a a a c c c a c a c g t a c c a g a t g a t g a a c 840
Db 840 A A G A C A C C T G C C C A C C A C T A T G T G T A C A C C C A C C A C C T A T G A T G A T G T C A A C 899
QY 841 c c g a g g a g a a t a g a c t t g t g c a c t g c g t g a a g a g t t c c c g t a a t t a t g t g 900
Db 900 C C G A A G G A G A G A C G T T T G G T G G C C A C C T G T G T A A A A G T G C C C C A A A C T A C G T G 959
QY 901 g t g a c a g a t c a e g t g t g c g t c c g a g c c t t g y g c c a g a c t a t a g a t g a g a g a 960
Db 960 G T A C A G A T C A G T A G C T A T G T G T C A G C C T G G G C T G A C T C T A C G A A G A G A G A A 1019
QY 961 g a c g c t c g c a a g t t a a g a a g t g c g a a g g c t t g c c g a a g t g t a a c g g a t a 1020
Db 1020 G A T G C A T C C G A G A G T A A A A A T G A T G G G C C T G C C A A G T T T A A T G C A T A 1079
QY 1021 g t a t g t g a a t t t a a g a c t c a c t c a t a a a t g c a a a t a t a a c t t a a a c t t c a a 1080
Db 1080 G C A T A T G G T G A A T T A A A G A C A C A C T C T C C A T A A T G C A A A C A T C A A A C A C T T C A A A 1139
QY 1081 a a t g a c c t c a t a t a g t g g a t c t c c a a t c c t g c c g t g a g a t t a g a g t g a c c 1140
Db 1140 T A C T G A C T G C A T A G G G G A C C T T C A C A T C T C C A G T G C C T T A A G G G A T T C T 1199
QY 1141 t t c a c a t a c t c c t c t c t g a t t c a c a g a a c t g a t a t c t g a a a c g t a a g a a 1200
Db 1200 T T C A C C C G C A C T C C T C T A G A C C C A C A G A A C T A A A A T T C A A A A C C G T A A G A A 1259
QY 1201 a t c a c a g a t t t g c t g a t t c a g g c t g a a a c a g a g a c c t c a t g c c t t 1260
Db 1260 A T A C A G C G C T T T T G C T G A T T C A G C C T G G C C G A T A C T G A C T G A C C A T G C T T T C 1319
QY 1261 g a g a a c c a g a a t a a g c g c g a g a c c a a g a a a t g t c a g t t t c t c t t g a g t c 1320
Db 1320 G A A A C T T A G A A A T A A T A C G T G A G A C A A G A A C A A C A T G C T T T T T G G C G G C T C 1379
QY 1321 g t c a g c t g a a c a t a a t c c t t g g a t t a c g t c c t c a a a g a g a t a a g t a t a g a t 1380
Db 1380 G T T G G C T G A A C A T C A C A T C A C T A C G G G C T G C T T C C C A A G G A G A T A G T A T G S G A T 1439
QY 1381 g t a a a a t t c a g a a c a a a a t t t g t c t a t a a a t a a a c t t g a a a a a c t g 1440
Db 1440 G T A T C A T T T C G A A A C C A A A T T T G T G T A C G A A A C A A A T A A A C T G A A A A A A C T C 1499
QY 1441 t t t g a c c t c g g t c a g a a a c c a a a t t a a g c a a g a g t g a a a c a g t g c a a g 1500
Db 1500 T T G G G A C A C C C A T C A G A A A C C A A A A T T A T G A A C A G A G T G A A A G A A G C T C A A G 1559
QY 1501 g c a c a g c a g a g t c t g c a t g c c t g t c c c c a g a g g c t g c t g g g c c c g a g a c c 1560
```


Dh	1120	GATGGGATCCGGAAGTGTAAAAAATGTGATGGGCGCTTGCCAAAGTTTGTATGGCATA	1179
Oy	1021	ggatcttggaaattttaaagactcaactctcataaattgtaagaaattaaactcttaaa	1080
Dh	1180	GGCATGGGGAAATTAAAGACACTCTCCATTAATAGCTACAAACATCAACACTTCACA	1239
Oy	1081	aacttgacctccatcaatgtagaattccaaactctgcgcgtgagcatltaagggtgactcc	1140
Dh	1240	TACTGCACAGCCCATCAGGGGGAGACCTTCCACATCTGCCAGCGGCGCTTTAAGGGGATTC	1299
Oy	1141	ttcaacaaactccctcccttgagatccaaaggaaactggaattcttgaanaacgtaaaagaa	1200
Dh	1300	TTGACCGCGACCTCCCTCTAGACCCACGGAACCTAGAAATTCTTAAACCGTAAGAGAA	1359
Oy	1201	atacaagggtttttgctgattcagcttgagcttgagaaacaggaagacctccatgctctt	1260
Dh	1360	ATTAACAGGCTTTTTCCTGATTTCAAGCTTGCGCTTGAACTGGACTGACCTCCATGCTTTC	1419
Oy	1261	gagaacactgaaataataacgagcgagagacaaagacaatgltcagtttctcttgagtc	1320
Dh	1420	GAGAACCTTGGAAATATATACGTGCGACAGCAAAAGCAAACTGCTCAGTTTCTTGGCGGAC	1479
Oy	1321	gtcaagctgaataaataacctcttgaggatctgactccctcaaggagataaattggaagt	1380
Dh	1480	GTTTGCCCTTAACATATACATCTACTGSGGGCTGCGTTCCCTCCAAAGAGATCATGATGGAGAT	1539
Oy	1381	gtgataaatttcagyaacaanaaattgtgctatgcaaatataaataaactgnaaaaactg	1440
Dh	1540	GTTGATATTTCCTCGGAACCCGAATTTGTGCTACGCAAAACACAAATTAACCTGGAAAAATC	1599
Oy	1441	tttggaacctccgycgaanaaaccaanaattataaagcaacagaggctgaanaacactgcaag	1500
Dh	1600	TTGCGAGACACCCAAATCAGAAAAACCAAAATGATGAACACAGCTGAGAAAGACTGCAG	1659
Oy	1501	ggcaacagcgcaagttgtgcagctgtgtgtcccgagagctctctgggcccggagccc	1560
Dh	1660	GCGGTACACACAGCTGTGCAATCTTTATGTCTCTCGAAGGCTGTGAGGCGCTTGAGCCC	1719
Oy	1561	agggaactgcgtctcttgccggaattgtcaagcggaagcaggaatgcytggacaagtgaag	1620
Dh	1720	AGGGACTGTGTCTCTCCGCAAAATGTGAGAGAGCGACGGAGTCCGTGGAGAAATGGAAC	1779
Oy	1621	ctctcggaagctgagcgaagggaatttgtgagaagacttgatgcatacagtgccaccca	1680
Dh	1780	ATCTCGAGGGGAGAACCAAGGAGATTGTGAAAAATTCGAATGCATTCACGTGCATCCA	1839
Oy	1681	gaatgctcctgcacgaagcatalaacttcaacttgacaagaagcggaacaaactgtatc	1740
Dh	1840	GAATGCTCCGCCAGGCGCATTAACCTGTACAGGAGAGGGGCGACGACAACTGCATC	1899
Oy	1741	caatgtgaccactataacttgaagcgcccccactgtgttaagaaactgcgccggagagatcatg	1800
Dh	1900	CAGGTGCGCCACTACATTGATGAGGCCCACTGTGCTAAGACTCTCCAGCTGGACATCAG	1959
Oy	1801	ggagaaacaacaacacttgattgaaataagcagaagccgcactgtgtgtccacactgtgc	1860
Dh	1960	GAGAGAGACAAACACTCTGTGTGGAATATGACGATGCCAATTAATATGCTGCGACCTATGC	2019
Oy	1861	catccaacttgacactcaagatgcaactgggcagagtccttgaagagctgtcca-----aag	1919
Dh	2020	CACCCCAACCTGTAACCTATGAGTGTGCTGGGCCACAGGCTTTCAAAGAGATGTGAATGTGGCCA	2079
Oy	1915	aatggctcaaaagatccgltccatcagccaacgaggaatggtgtggggccctctgtgcgtgtg	1970
Dh	2080	TCGTGGCGCAAAATATCCATTATATGCGCACACGGGATGTGGGGGCGCTCTCTTCATATGTTG	2139
Oy	1975	gtgtgtgcccctggagatcgactcttcaatgcgaaggcgcaacatggttgtgaaggcgac	2033
Dh	2140	GTTGTGGCGCTTGTGGATTTGGCTGTATGATGCGAAGACGTCAATATGTTTCCAAAGCGTAC	2198

[illegible]

QY 1501 gccacagcagctctgcatgctctgctcccccagagcctgctgagccgagccc 1560
DB 1715 gccctgaacacacctctgcaatcccttgcgcgaagccctgagagccgagccc 1774
QY 1561 agggagctgctctctgctcagatgctcagcagcagcagcagcagcagcagc 1620
DB 1775 agggagctgctctctgctcagatgctcagcagcagcagcagcagcagcagc 1834
QY 1621 cttctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1680
DB 1835 atctctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1894
QY 1681 gagtctgctcctcagcagcagcagcagcagcagcagcagcagcagcagcagc 1740
DB 1895 gaattgctctcctcagcagcagcagcagcagcagcagcagcagcagcagcagc 1954
QY 1741 caagtctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1800
DB 1955 caagtctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2014
QY 1801 gggag 1860
DB 2015 gggag 2074
QY 1861 catcacaactgac 1908
DB 2075 caccac 2122

RESULT 13
LOCUS HSU48722 1593 bp mRNA PRI 05-NOV-1996
DEFINITION Human epidermal growth factor receptor precursor (EGFR) mRNA,
complete cds.
VERSION U48722.1 GI:1628549
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Reiter, J.L. and Mailhe, N.J.
TITLE A 1.8 kb alternative transcript from the human epidermal growth
factor receptor gene encodes a truncated form of the receptor
Nucleic Acids Res. 24 (20), 4050-4056 (1996)
JOURNAL 97078686
MEDLINE 2 (bases 1 to 1593)
REFERENCE Reiter, J.L.
AUTHORS Direct Submission
TITLE Submitted (07-FEB-1996) Jill L. Reiter, Biochemistry and Molecular
Biology, Mayo Clinic, 200 First St. SW, Rochester, MN 55905, USA
LOCATION/Qualifiers

FEATURES

source
1. 1593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p11.2-p12"
/tissue="type-placenta"
1. 1456
/gene="EGFR"
1. 244
/gene="EGFR"
245. 1462
/note="ErbB-1"
/product="epidermal growth factor receptor precursor"
/protein_id="A450802.1"
/db_xref="GI:1628550"
/translacion="KRPSTGAGAAIILALALCAPSALFEKKVCCGTSNKLITQICTP
EDHFLSLDRNFNCEVAVLGNLEITVGRNVLSTLKTIOVAVYVILALNTERIPLE
NLOITRGNMYENSVYLAVALSNYDANKTGLKELPMRLDIEILGAVRFSNNPDLNCE

STORIDVSDPLSNMSMDPQNHGSCOKDPCSPGSCGAGSENCOKLITKICAOQ
CSGCRKRSFSDCHNOCAAGCTGFRSDCLVCRKRIDEATCKDKCPMLYINPTTQA
MDVNPBGKYSFSGATCYKCKPCRYVVTGHSVACACADYEMEBGVRCKKCGPGR
KVCNIOIGEFKDSLSINATNLIKFKMCTSISSDLILVAFRSDSFTHPTDPOEL
DLKTVKEITGLS"
245. 316
/gene="EGFR"
317. 1459
/product="epidermal growth factor receptor"
626. 634
/gene="EGFR"
/note="encodes Asn-linked glycosylation site"
767. 775
/gene="EGFR"
/note="encodes Asn-linked glycosylation site"
830. 838
/gene="EGFR"
/note="encodes Asn-linked glycosylation site"
1055. 1063
/gene="EGFR"
/note="encodes Asn-linked glycosylation site"
1298. 1306
/gene="EGFR"
/note="encodes Asn-linked glycosylation site"
1325. 1333
/gene="EGFR"
/note="encodes Asn-linked glycosylation site"
1452. 1456
/gene="EGFR"
/note="5' splice donor site"
1463. 1569
polya_signal
1537. 1542
polya_site
1569 a 445 c 424 g 318 t
BASE COUNT 406 a 445 c 424 g 318 t
ORIGIN

Query Match 59.4% Score 1207.8; DB 9; Length 1593;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1209; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atggagacccctcggagagagcggcggcggcggcggcggcggcggcggcggcggcggc 60
DB 245 atggagacccctcggagagcggcggcggcggcggcggcggcggcggcggcggcggc 304
QY 61 gggagctgggctctgggagagagagagagagagagagagagagagagagagagag 120
DB 305 gggagctgggctctgggagagagagagagagagagagagagagagagagagagag 364
QY 121 ctgggacactcttgagagatctctcagcctcagagagagagagagagagagagagag 180
DB 365 ttggggacacttttgagagatctctcagcctcagagagagagagagagagagagagag 424
QY 181 gtccttgag 240
DB 425 gtccttgag 484
QY 241 accatccag 300
DB 485 accatccag 544
QY 301 ctggagaaacctgagagac 360
DB 545 ttggagaaacctgagagac 604
QY 361 gtcctatctaaactatgag 420
DB 605 gtcctatctaaactatgag 664
QY 421 caggaacccctgag 480
DB 665 caggaacccctgag 724

Oy 481 agatccagctgagcgagacatagtcagcaagtacgtcttcacgaacatgctgagatgac 540
 |||||||
 Db 725 AGCATCCAGTGGGGGACATAGTACGACACTGACTTCTTCACACCAACATGTCATGAGACTTC 784
 Oy 541 cagaacacactgagcgagctgacccaagaatgtatgacaaagctgtcccaatggaagctgtg 600
 |||||||
 Db 785 CAGAACCACTGGGACGCTGCCAAAAGTGTGATCCAACTCTCCCATGGAGAGCTGCTGG 844
 Oy 601 ggtgcaagagagagagacatgacgaacaaactgacaaatcatctgtgtcccgacagctgtcc 660
 |||||||
 Db 845 GGTGACAG 904
 Oy 661 gggcgctgctgagcaagctgaccccaagctgacgtgctgacaaacagctgtgtgctgagctg 720
 |||||||
 Db 905 GGGCGCTGCGCTGGCAAGTCCCGACAGTACGCTGCCACAAACAGTGTGTCGACGCTGC 964
 Oy 721 acagagcccgag 780
 |||||||
 Db 965 ACAGGCCCCGGGAG 1024
 Oy 781 aagagacacgtgcccccaactcaltgctctcaaaaccccaacacgtaaccagatgagatgagac 840
 |||||||
 Db 1025 AAGGACACCTGCCCCCCTCCTCATGCTCTCTACAAACCCACACAGTACAGAGATGATGATAC 1084
 Oy 841 cccgaagagcaatacagctgtgtgacccctgctgagaaagctgtcccgatgagatgag 900
 |||||||
 Db 1085 CCGGAGGGCAAAATACAGCTTTGGTGCGACCTGCTGAGAGAAAGTGTCCCGTAATTTATGTG 1144
 Oy 901 gtgacagatacagagctgtgtgctgacagagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960
 |||||||
 Db 1145 GTGACAAATACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1204
 Oy 961 gacgagctgagcaagctgt 1020
 |||||||
 Db 1205 GAGGCGCTCGCGCAAGTGAAGAGTGCAGAGGCGCTTGCAGCAAGTGTGAACGAGATA 1264
 Oy 1021 ggtattgtgtatatttaagaactcactcctcctcaaaatgctgaatataataaacttcaaa 1080
 |||||||
 Db 1265 GGTATGTTGTAATTTAAAGCTACCTCTCATTAATTAATTAATTAATTAATTAATTAATTAAT 1324
 Oy 1081 aactgacacctcactcagctgt 1140
 |||||||
 Db 1325 AACTGACCTCTCATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
 Oy 1141 ttcaacatactctctctctgt 1200
 |||||||
 Db 1385 TTCAACATTACT 1444
 Oy 1201 atcacaaggtt 1211
 |||||||
 Db 1445 ATCACAGTTT 1455

RESULT 14
 LOCUS 180039 1868 bp DNA PAT 10-JUN-1998
 DEFINITION Sequence 1 from patent US 5708156.
 ACCESSION 180039
 VERSION 180039.1 GI:3208329
 KEYWORDS
 SOURCE
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1868)
 AUTHORS Alexs,V.
 TITLE Epidermal growth factor receptor-like gene product and its uses
 JOURNAL Patent: US 5708156-A 1 13-JAN-1998;
 FEATURES
 source 1..1868
 1..1868 /organism="unknown"
 BASE COUNT 449 a 554 c 484 g 371 t
 ORIGIN

Query Match 59.4%; Score 1207.8; DB 6; Length 1868;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1209; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 atgagacacctcggagacggcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 60
 |||||||
 Db 504 ATGCGACCTCTCGGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 563
 Oy 61 gcgagctggagcttggagagaaagaaatlttgcaagggcagcgagaaagcaagctcagcgag 120
 |||||||
 Db 564 GCGAGTGGGCTCTGGAGGAGAAAGAAAGTTTCCAAAGGACAGATTAACAAGCTACAGCGAG 623
 Oy 121 ttgggcaacttgaagatcatttctcagccctcagagagatgtcaactgtgagagtg 180
 |||||||
 Db 624 TTGGGCACTTTTAAATATCATTTTCTCAOCCCTCCAGAGATGTTTCAATACCTGTAGAGTG 683
 Oy 181 gtccctggagaaattggaatltacatagtgtgagagaaatltgacttccctccttaag 240
 |||||||
 Db 684 GTCCCTGGAAATTTGAAATTAACCTATGTGAGAGAAATTTATGATCTTCTTAAAG 743
 Oy 241 accatccagagagtgagctgtgtatgttccctcattgtccctcacaacagtgagcgaaatccct 300
 |||||||
 Db 744 ACATTCAGAGAGTGCGCTGTTATGCTCATGTCCTTCACACAGTGGAGCGAATTCCT 803
 Oy 301 ttggaacacctgcaagatcactcagagaaatgtactcgaatattcctatgcttaga 360
 |||||||
 Db 804 TTGGAACACCTGCAATCATCAGAGAAATATGTAAGAAATTTCTTATGCTTAGCA 863
 Oy 361 gtcttactaactatgatagtcaataaaccggactgaaagagctgtgcccataagaaatltta 420
 |||||||
 Db 864 GTCTTATCTAATCATATGATCAATTAACCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 923
 Oy 421 caggaatactcctgtagcgccgt 480
 |||||||
 Db 924 CAGGAATCTCTGATGCGCGCGCTGCGGCTTACAGAACACCTGCGCTGTGCANCTGGAG 983
 Oy 481 agatccagctgagcgagacatagtcagcaagtacgtcttcacgaacatgctgagatgac 540
 |||||||
 Db 984 AGCATCCAGTGGCGGAGACATAGTACAGCACTGACTTTCACACCAACATGTCATGAGACTTC 1043
 Oy 541 cagaacacactgagcgagctgtgcaaaagtgtgacaaagctgtgtcccaatggaagctgtg 600
 |||||||
 Db 1044 CAGAACCACTGGGACGCTGCCAAAAGTGTGATCCAACTGTCCCAATGGAGAGCTGCTGG 1103
 Oy 601 ggtgcaagagagagagacatgcccagaacactgacaaatcatctgtgtcccgagctgtcc 660
 |||||||
 Db 1104 GGTCCAGGAG 1163
 Oy 661 gggcgctgagctgtgcaagctcaccagctgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
 |||||||
 Db 1164 GGGCGCTGCGCTGGCAAGTGTCCCGAGTGTGCTGTGCTGCACAAACAGTGTGCTGCCAGGCTGC 1223
 Oy 721 acagagcccgag 780
 |||||||
 Db 1224 ACAGGCCCGCGGAG 1283
 Oy 781 aaggaacacctgcccccaactcaltgctctcaaaaccccaacagtaaccagatgagatgagac 840
 |||||||
 Db 1284 AAGGACACCTGCCCCCCTCCTCATGCTCTCTCAACCCACACAGTACAGATGAGATGAGAC 1343
 Oy 841 cccgaagagcaatacagctgt 900
 |||||||
 Db 1344 CCGGAGGGCAAAATACAGCTTTGGTGCGACCTCTGCGAAGAAAGTGTCCCGTAATTTATGTC 1403
 Oy 901 gtgacagatacagagctgt 960
 |||||||
 Db 1404 GTGACAAATACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1463
 Oy 961 gacgagctgagcaagctgt 1020
 |||||||
 Db 1464 GAGGCGCTCGCAAGTGTAGAGAGTGCAGAGGCGCTTCCCGCAAGTGTGTACGGAGATA 1523


```
Db 1404 GTGACAGATCACGGCTCGTGGCTCCGAGCCTGTGGGGCCGACAGCTATGAGATGAGGAA 1463
QY 961 gacggcgtccgaagtgtaagaagtgcgaaggcccttgccgcaaatgltgaaggaata 1020
    |||||
Db 1464 GACGGCGTCCGCAAGCTAAGAGTGGCAAGCGCCCTGCCGCAAAAGTGTGTATACGGATA 1523
    |||||
QY 1021 ggtattggtgaatttaagaactcactctccataatgctacgaatatlaaacactcaaa 1080
    |||||
Db 1524 GGTATTGGTGAATTAAAGACTCACTCTCCATTAATGCTACGAATATTAACACTTCAAA 1583
    |||||
QY 1081 aactgacctccatcaagtcgagatccacalccctgcggtgycatlltaggggtgactcc 1140
    |||||
Db 1584 AACTGCACCTCCATCAGTGGCGATCTCCACATCCTGCCGGGTGGAATTTAGGGGTGACTCC 1643
    |||||
QY 1141 ttacacatactccctcctctgatatccacagaactgatatlcgaaaaccgtaagaaa 1200
    |||||
Db 1644 TTCACACATACCTCCCTCTGTGATCCACAGAACCTGATATCTGAAAAACCGTAAAGGAA 1703
    |||||
QY 1201 atcacagggtc 1211
    |||||
    ||
    ||
Db 1704 ATCACAGGTTT 1714
```

Search completed: January 14, 2002, 20:13:47
Job time: 4172 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 14, 2002, 19:05:00 ; Search time 159.75 Seconds

(without alignments)
10915.803 Million cell updates/sec

Title: US-09-715-249-1_COPY_1_2034

Perfect score: 2034
Sequence: 1 atgcgaccctccggagcgc.....acatcgttcgagacgacacg 2034

Scoring table: IDENTIFY_MUC
Gapop 10.0 ; Gapext 1.0

Searched: 930621 seqs, 428662619 residues

1 number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

N_Geneseq_1101:*

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2034	100.0	3633	22	AAF85332
2	2034	100.0	3818	21	AA661552
3	2034	100.0	5532	21	AA23954
4	1572.4	77.3	6274	14	AAQ43812
5	1425.8	70.1	6207	14	AAQ43813
6	1207.8	59.4	1868	19	AAV06408
7	992	48.8	1958	22	AA84211
8	981	48.2	5750	14	AAQ43814
9	431	21.2	5484	22	AAH47973
10	431	21.2	5501	15	AAQ64896
11	431	21.2	5501	17	AA18532

12	431	21.2	5555	15	AAQ64898	HER4 with alternat
13	431	21.2	5555	17	AA18533	Receptor tyrosine
14	419	20.6	3768	21	AAA09455	Human heregulin 2
15	419	20.6	4530	16	AA701595	Her-2/neu (ERBB2/c
16	419	20.6	4530	18	AA71253	Human HER2 gene.
17	419	20.6	4530	21	AA260815	Nucleotide sequenc
18	419	20.6	9274	22	AA24257	HER2 transgene pla
19	417.4	20.5	3500	21	AA689736	Human HER-2/neu co
20	417.4	20.5	3768	17	AA740739	HER-2/neu oncogene
21	417.4	20.5	3768	17	AA740739	Human HER-2/neu on
22	417.4	20.5	3768	22	AAH23392	Sequence encoding
23	417.4	20.5	4299	14	AAQ46083	cdNA encoding the
24	417.4	20.5	4472	21	AAA14812	HER-2 nucleic acid
25	417.4	20.5	4473	20	AA231071	HER-2 nucleic acid
26	409.4	20.1	4805	17	AA100059	erbB-3 cDNA clone
27	409.4	20.1	4805	19	AAV58734	erbB-3 glycoprotein
28	409.4	20.1	4805	20	AA660313	E3-16 cDNA clone e
29	406.2	20.0	4978	12	AAQ13363	HER-3 epithelial g
30	402.2	19.8	3955	16	AA701590	Rat neu promoter.
31	402.2	19.8	3955	21	AA689753	Rat HER-2/neu prot
32	400.6	19.7	2385	18	AA72725	Her-2-GM-CSF immuno
33	400.6	19.7	2871	21	AA250586	DC8scv-erbB2EC fu
34	395.8	19.5	3771	22	AA250586	Mouse Her-2/neu cd
35	395.8	19.5	3771	22	AAH42210	Nucleotide sequenc
36	392	19.3	4545	12	AAQ12224	EGFR-R erbB-3 gene
37	391	19.2	1872	11	AAQ06828	Extracellular port
38	172	8.5	172	21	AAQ61547	DNA fragment of a
39	149	7.3	1033	21	AAQ61546	DNA fragment of a
40	137	6.7	220	22	AA120955	Probe #10888 for g
41	137	6.7	220	22	AA146199	Probe #14888 used
42	137	6.7	220	22	AA106657	Probe #6658 used t
43	137	6.7	486	22	AA11743	Probe #1676 for ge
44	137	6.7	486	22	AA13050	Probe #1736 used t
45	137	6.7	486	22	AA101671	Probe #1662 used t

ALIGNMENTS

RESULT 1

ID AAF85332 standard; DNA: 3633 BP.

AC AAF85332:

23-JUL-2001 (first entry)

Nucleotide sequence of wild type EGFR1.

Epidermal growth factor receptor 1; EGFR1; cell surface marker;

KW protein-tyrosine kinase receptor; PTKR; immunoselection; gene therapy;

KW bone marrow transplant; graft facilitation; immune reconstitution; ss.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

MO200136659-A2.

25-MAY-2001.

17-NOV-2000: 2000NO-EP11474.

19-NOV-1999: 9905-0166594.

19-NOV-1999: 9905-0444038.

30-MAR-2000: 200005-0539248.

(NOVS) NOVARTIS AG.

(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

Location/Qualifiers

I..3633

/*tag= a

/transl_except= "(pos: 1618..1620, aa: Asn)"

/product= "EGFR1"

QY 121 ttggacatttgaagatcatttctcagctccagagatgttcaatactgtgagtg 180
 |||
 Db 307 ttggacatttgaagatcatttctcagctccagagatgttcaatactgtgagtg 366
 QY 181 gtccctgggaatttgaatacttactatctgacagaggaattatgactctctctaaag 240
 |||
 Db 367 gtccctgggaatttgaatacttactatctgacagaggaattatgactctctctaaag 426
 QY 241 accatccagagagtggtcgtgtatctcattctccatcaacacagtgagacgattctc 300
 |||
 Db 427 accatccagagagtggtcgtgtatctcattctccatcaacacagtgagacgattctc 486
 QY 301 ttggaacccctgcagatccatccagaggaataatgtactcgaataattctatgtcttaga 360
 |||
 Db 487 ttggaacccctgcagatccatccagaggaataatgtactcgaataattctatgtcttaga 546
 QY 361 gtcttactcaactatgacataataaaccggaactgaagagctctccatgaagaaatcta 420
 |||
 Db 547 gtcttactcaactatgacataataaaccggaactgaagagctctccatgaagaaatcta 606
 QY 421 caggaatactctcagatccagagctgtcgtgtatcagacacacccctgtctgacagtgag 480
 |||
 Db 607 caggaatactctcagatccagagctgtcgtgtatcagacacacccctgtctgacagtgag 666
 QY 481 agcatccagatgtgcgggaatactgacagatgtacttctccagacacatgtctgagacttc 540
 |||
 Db 667 agcatccagatgtgcgggaatactgacagatgtacttctccagacacatgtctgagacttc 726
 QY 541 cagacacacccctgcagatccagacacacacacacacacacacacacacacacacacacac 600
 |||
 Db 727 cagacacacccctgcagatccagacacacacacacacacacacacacacacacacacacac 786
 QY 601 ggttcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 660
 |||
 Db 787 ggttcaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 846
 QY 661 ggtgcgtctgcgtgtgcagatgtcccccagtgactgtctgcacacacacacacacacacacac 720
 |||
 Db 847 ggtgcgtctgcgtgtgcagatgtcccccagtgactgtctgcacacacacacacacacacacac 906
 QY 721 acaagccccctgcag 780
 |||
 Db 907 acaagccccctgcag 966
 QY 781 aaggaacacccctgcagac 840
 |||
 Db 967 aaggaacacccctgcagac 1026
 QY 841 cccgagaggaataaagcttctgtgcacacacacacacacacacacacacacacacacacacac 900
 |||
 Db 1027 cccgagaggaataaagcttctgtgcacacacacacacacacacacacacacacacacacacac 986
 QY 901 gtcgac 960
 |||
 Db 1087 gtcgac 1146
 QY 961 gacgggctcccgagaggtgtgaagaagtgagagagagagagagagagagagagagagagagag 1020
 |||
 Db 1147 gacgggctcccgagaggtgtgaagaagtgagagagagagagagagagagagagagagagagag 1206
 QY 1021 ggtatctgtgtgaatttgaagaactcactcactcactcactcactcactcactcactcactc 1080
 |||
 Db 1207 ggtatctgtgtgaatttgaagaactcactcactcactcactcactcactcactcactcactc 1266
 QY 1081 aactgcacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1140
 |||
 Db 1267 aactgcacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1326
 QY 1141 ttccacatactcactcactcactcactcactcactcactcactcactcactcactcactcactcactc 1200
 |||
 Db 1327 ttccacatactcactcactcactcactcactcactcactcactcactcactcactcactcactcactc 1386
 QY 1201 atccacaggttcttctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1260

Db 1387 atccacaggttcttctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1446
 QY 1261 gagaacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1320
 |||
 Db 1447 gagaacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1506
 QY 1321 gtcacacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1380
 |||
 Db 1507 gtcacacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1566
 QY 1381 gtcacacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1440
 |||
 Db 1567 gtcacacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1626
 QY 1441 ttctggacccctcctgcagaggaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1500
 |||
 Db 1627 ttctggacccctcctgcagaggaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1686
 QY 1501 gtcacacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1560
 |||
 Db 1687 gtcacacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1746
 QY 1561 agggacacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1620
 |||
 Db 1747 agggacacccctcactcactcactcactcactcactcactcactcactcactcactcactcactc 1806
 QY 1621 ctctctggaggttgagccag 1680
 |||
 Db 1807 ctctctggaggttgagccag 1866
 QY 1681 ggtgtgcctcctcag 1740
 |||
 Db 1867 ggtgtgcctcctcag 1926
 QY 1741 cagtggtgcctcctcag 1800
 |||
 Db 1927 cagtggtgcctcctcag 1986
 QY 1801 ggaagaaac 1860
 |||
 Db 1987 ggaagaaac 2046
 QY 1861 catccaaacttgac 1920
 |||
 Db 2047 catccaaacttgac 2106
 QY 1921 cctaaagatccgttccatctgcac 1980
 |||
 Db 2107 cctaaagatccgttccatctgcac 2166
 QY 1981 gcccttgagatccgttccatctgcacacacacacacacacacacacacacacacacacacac 2034
 |||
 Db 2167 gcccttgagatccgttccatctgcacacacacacacacacacacacacacacacacacacac 2220

RESULT 4
 ID AAO43812 standard; DNA: 6274 BP.
 AC AAO43812;
 DT 20-OCT-1993 (First entry)
 DE pRLID2D3, Apal construct.
 XX
 KW Epidermal growth factor receptor truncate protein; EGF; growth;
 KM binding sites; adsorptive agents; mammalian cell growth abnormality;
 OS detection; reproduction; signal transmission; pRLID2D3Apal; ds-
 XX Synthetic.
 FH
 Key Location/Qualifiers

RESULT	6	
AAV06408	standard; DNA; 1868 BP.	
AAV06408		
AAV06408;		
05-MAY-1998	(first entry)	
Epidermal growth factor receptor-like protein (TEGFR) encoding CDNA.		
Epidermal growth factor receptor-like protein; TEGFR; human; antibody		
diagnosis: cancer; hyperproliferative disease; ss.		
homo sapiens.		
Key	location/Qualifiers	
CDS	504..1721	
	/*tag= a	
	/product= TEGFR	
	504..575	
	/*tag= b	
	/note= "putative signal peptide"	
	576..1718	
	/*tag= c	
US5708156-A.		
13-JAN-1998.		
31-MAY-1996;	96US-0658883.	
31-MAY-1996;	96US-0658883.	
ILEK/ ILEKIS J V.		
ILEKIS JV;		

CC This cDNA encodes a novel human epidermal growth factor receptor-like
CC protein termed TEGFR. TEGFR is derived from human placenta. Fragments of
CC the TEGFR cDNA can be used as a probe in a hybridisation assay to detect
CC full length of TEGFR. An antibody can be produced using the antigen
CC comprising at least the unique C-terminal sequence of the TEGFR
CC polypeptide, optionally conjugated to an immunogenic carrier. The
CC antibody can be used in an immunoassay to detect the polypeptide. The
CC hybridisation assay and other inhibitors can be used in the diagnosis of human
CC cancers. Antihodlers and other inhibitors of the TEGFR polypeptide may
CC also be used to treat hyperproliferative diseases including cancer.

Query Match	59.48;	Score 1207.8;	DB 19;	Length 1868;
Best Local Similarity	99.8%;	Pred. NO. 0;		
Matches 1209;	Conservative	0;	Mismatches 2;	Indels 0;
				Gaps 0.

[illegible]

Dh	767	aacgctcttcgacgacatgaaattccaaatgaaacagctgagacatctgtcttaccatgctcccaaa	826
Oy	800	tcagcttcacaaaccccccgcgtacacagatgaaatgaaatgaaaccccagaaagaaataagct	859
Dh	827	cccttcgtcccaaaaccacacccccttcacactgagagacacaaatttcaatgcaaaatgaaacat	886
Oy	860	ctggctgcacactgcgtctgaaagaagatgctcccgcaactatctatgctgaaacagatcacagctctg	919
Dh	887	atcgagacatctccgtgctcaaaagaaatgctcacaaataaactcttggtc---agattccagattcttc	943
Oy	920	gcgtccgcagacactgctggggccgacacgtacatgaaatgaaagaaagaaacagcgtccgcgaatgcta	979
Dh	944	gtctgtcgtctccgcccctcagcttcacaaagatgaaagaaagaaagaaagaaatgtaaaatgtgcta	1003
Oy	980	agaaagctgcgaagagcctctgcgcgcgaatgctgtaacgaaatgaaatctgaaatctgaaag	1039
Dh	1004	aaactctgcactgcacatctgtcccaaaagctctgaaatgagcattgcgtcacacagatctatgtgt	1069
Oy	1040	aaatcacatcccaataaaatgtctcgtgaaatgttaaaagacatcaaaataatgcacatcccatcaag	1099
Dh	1064	cagctccagacctgctgattccacgtataacatgacaaataccaaatcaacatgcaacaaatccaaatg	1122
Oy	1100	gcgaatcccaacatccctgcacgctgagcatcttaaggggtgaaatcccttcacacatacctcctctc	1155
Dh	1124	ggaaatttgattctctcattgatactcgtgtaattctaaatgaaagaccccttacaatgtaaaatgaaagca	1183
Oy	1160	tgaataccaaagaaatctgaaatatactctgcgaaaacccgtcaaaagaaataacagaaagatttttgtga	1239
Dh	1184	taagacccaagaaagaaatctgaacgctctcttcgtgcacagctcaagaaatgaacaaagctctctctgaaca	1244
Oy	1220	ttcaagcctctgcagctctgaacaaacagacagacacacacacacacacacacacacacacacacacac	1279
Dh	1244	taagatcagatgcacacaaacaaacaaatgacatgcagctcagctgtttttcttaacacctgtgacacattg	1304
Oy	1280	ggcgacagagacaaagaaacaaatgctaaagtctctctctgcagctctgcgtacagacctgaaataaact	1339
Dh	1304	gtggaaagaaatcactcaatctatgctgctgcctctgctatactccacagaaacagtgacatacact	1365
Oy	1340	cccttgagattacgctctccctccaaagaaatgaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg	1399
Dh	1364	ctctcaacaaatctcagagacccctctggaaagaaatctcagctgcagaaacatactataactctgaacaa	1423
Oy	1400	aaagattctgtcctatgacaaatacaatacaatacaatacaatacaatacaatacaatacaatacaatacaata	1455
Dh	1424	gcaaacctctgtctataataatacaatacaatacaatacaatacaatacaatacaatacaatacaatacaata	1483
Oy	1460	aaacacaaatactaaatgaacaaacagaaatgtaaataacagctgcgaagaaacagacagacagacagac	1519
Dh	1484	gaattaaatctaaatccggaacacagaaagaaagaaatgtctaaatctgaaagaaatgctgta	1543
Oy	1520	aaatgctctgtgcctcccccagaaagagctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcctgcct	1579
Dh	1544	aaacatctgctgtctcagctgaaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1603
Oy	1580	gaaatctcagacagaaagctgaaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1638
Dh	1604	gacgctctcagatgaagaaagaaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1663
Oy	1640	ggagaaatcttgctgaaagaaatctcagatgtgcaatacaatgctccacacagaaagtg---cgcgtccacag	1696
Dh	1664	ggagaaatcttgctgaaagaaatgctcacaatctcagatgtgaaatgtgcaaccccagatgctgaaagaaatgaaag	1723
Oy	1697	cgaatgaatacaatacaatacaatacaatacaatacaatacaatacaatacaatacaatacaatacaatacaata	1755
Dh	1724	aaatgctctcctcaatgacaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg	1783
Oy	1757	ttgaacgctccccaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1816
Dh	1784	aaagaaatcccaaaacagctgctgaaataaaatgctccaaagaaatgctccaaagaaatgctccaaagaaatg	1843
Oy	1817	tgatctgaaagaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaatg	1876
Dh	1844	ttttc---aaagaaatgctgaaatcccaatgctgaaatgctgaaatgctgaaatgctgaaatgctgaaatg	1900

Oy	1877	acgcatgcactggcgccaggtccttgaagctg	19078
Db	1901	aagggtgtaacggtgcccaatgatcatgactg	19311

RESULT 10

AAQ64896 standard; cDNA; 5501 BP.

AC AAQ64896;

11-JAN-1995 (first entry)

AA
DE
HER4.

XX	Human epidermal growth factor
KW	

KW cancer; neuronal tissue; m
KW carcinoma; primer; probe;

XX Homo sapiens.
OS

XX	Key	Location/Comments
FH		

FT	34.,3960
CDS	/*tag= a
FT	

XX
PN
FD500374-2

FM EFCJ2214-A.
XX
01-YUN-1004
PD

PD 01-JUN-1994.
XX

PE 23-NOV-1993; 93EP-011883
XX

PR 24-NOV-1992; 9205-098116
XX

PA (BRIM) BRISTOL-MYERS SQUI
XX

PI Culouscou J, Plowman GD,
XX

DR WPI: 1994-169599/21.
DR P-PSDB: AAR54841

XX New recombinant nucleic acid

tyrosine kinase expressed

PT
vectors, antibodies, ligands
PT
cancers

XX	Claim 4; Fig 1; 104pp; Eng
PS	

XX
CC HER4 is the fourth member

CC is expressed in some human
CC or muscle origin. HER4 po

CC In assays (e.g. of HER4 mRNA
CC breast carcinoma) and as p

XX
50
Sequence 5501 BP: 1605 A:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Query Match	21.28
Post-processed	24.18

Best Local Similarity 34.18
Matches 991; Conservative

QY 83 agaaagttctgccaagcgca

Db 110 agtcagtggtgcaggaacgg

QY 143 ttctcaqccctccaqaqatqt

db 170 accgaacctgacgaagta

203 cctatgtccagaggaattatct

27	ccccccccc	ccccccccc	ccccccccc
28			
29	ccccccccc	ccccccccc	ccccccccc

LD 230 LC897C6ALF90A9UCL6A6CL6999

```

Oy      1877   acgatgacctggcagcgacttgaaggctg 1907
          |  |||  |||  |||  |||  |||
Db       1901   aagggttgtaacgcgccactagtatgctg 1931

RESULT# 10
AAQ64896 standard; cDNA; 5501 BP.
ID     AAQ64896;
AC      AAO64896;
XX      11-JAN-1995 (first entry)
DE      HER4.
KW      Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;
        cancer; neuronal tissue; muscle tissue; neoplasm;
        carcinoma; primer; probe; PCR; ss.
OS      Homo sapiens.
Key     Location/Qualifiers
FT      34..3960
CD      /*tag= a

EP599274-A.
PD      01-JUN-1994.
XX      23-NOV-1993;    93EP-0118837.
PF      XX
PR      24-NOV-1992;    92US-0981165.
PA      (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI      Culousscou J, Ploewman GD, Shoyab M;
DR      WPT: 1994-169599/21.
P-PSDB: AAR54841.

New recombinant nucleic acid expressing HER4 - a new receptor
for tyrosine kinase expressed in some cancer cells, and related
vectors, antibodies, ligands etc, for diagnosis and treatment of
cancers

Citation 4 : Fig 1; 104pp; English.

HER4 is the fourth member of the EGFR-family of tyrosine kinases and
is expressed in some human cancers and in some tissues of neuronal
or muscle origin. HER4 polynucleotides, opt. labelled, are useful
in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp.
breast carcinoma) and as primers in PCR or as probes.

Sequence 5501 BP; 1605 A; 1173 C; 1230 G; 1493 T; 0 other:

Query Match              21.2% Score 431; DB 15; Length 5501;
Best Local Similarity 54.1%; Pred. No. 5; 9e-113;
Matches 991; Conservative 0; Mismatches 825; Indels 15; Gaps 5;

Oy      83   agaaatttcgaagcacgacgtatacaaacgcccacgacgactttggaccactttgaagatcat 142
          |||||  |||||  |||||  |||||  |||||  |||||
Db      110   agtcagtgltgtgacgagaacggagaataaatagctctctctcgtaccttgaaacagagt 169

Oy      143   ttctcagccctcagagagatgtccaataactctgtgagtggtgcttgtggaatttgaaaatta 202
          |||||  |||||  |||||  |||||  |||||  |||||
Db      170   accgagctctgcgaagtlactatgaataactctgtgagtggtgtcctaattgacaacttgagataa 229

Oy      203   ctatatgtcgacagaggaattatgatcttctctctttaaagacatccagaggtggtcgtgt 262
          |||||  |||||  |||||  |||||  |||||  |||||
Db      230   ccagatttcagacaacacggagacctctctcttcctcgtgcgtctgttcgaaagattacaagtct 289
```


Oy	83	agaaagaaatgacaaagacccagagaaacaaagatccacgaattgagaaatttgaaagatcatt	142
Db	110	agaaacagatgagacaaagaaacagaaatacaaaacagacacacacacacacacacacacac	169
Oy	143	lctccagccctcccaagagaaatgltcaaaatactgagagcagcctctggagaaattgaaatta	202
Db	170	accagagacctctgacaaagaaatacaaaagaaacacatgagagatgtgtaagagcaaacctgagagataa	229
Oy	203	cccaatgacaaaggaatataatgaatcctctccctctcaaaagacacacaaagaaagatgacgtgtc	262
Db	230	ccaaagatcgagacaaacacccgagac	289
Oy	263	agatcccatatgacctcaaaacagatgagagcgaattccctcttggaaaacccgagaaatacata	322
Db	290	accgtgtctagtgacctcaaaatacagttcgttaaccctccctctgagaaattacagcatattacc	349
Oy	323	gaagaaataatgtaactaaagaaataatctcaatgcatgacagctgttctcaaaatacgaatgcaaa	382
Db	350	gtggagacaaaaacttataatgagatcagataatgaccttgagcaataatttttaaatataagaaag	409
Oy	383	ataaaagac---gagactgaaagagagtgctcccaatgaagaattatcaagaaatactgcatgag	439
Db	410	atgaaagaaaccttgagaaacttcaaaagaaactggaataaaagaaactgcaagaataacccttaaaatggtg	469
Oy	440	ccagtcgaggttcaagcaaaacaaacccctgacacgtgagaaagatgagaaatgacaaatgagtcggagaa	499
Db	470	gaatctaaatgaaagaaacaaataatccctctgttataagcaagacaacattcatattgcaagata	529
Oy	500	tgaatcagacagatgaaacttctcagaagaaatgcaatgaaatccaaagaaacaaacacatgagtcagac	559
Db	530	tgtctcggaacaacaaatgacctctcaaaatctgaaactctgtgtcaaaataatggtatctcaagat	589
Oy	560	gccaataagatgtaattccaaagacgtgtccaaatggagacatgacatgagcgtgtgcaagaaagaaacac	619
Db	590	gtgagacgtgtgccaataaaatccatgt---actgagccgtgtgtctgagggacccaagaaataatcatat	646
Oy	620	gccaataaactgaacaaataatcatctctgtgcccacgaatgtctccgagcgtgtgctgagcgtgaagt	679
Db	647	gccaagacttgaacaaagagcgtgtgtgagcaagaaataatgtaagcgtgaagtgtctaaagacact	706
Oy	680	ccccaagatgaaatctgtcccaaaacaaatgtaatgctgacagcctgcaagaccccccagagaaagacgc	739

[illegible]

Qy	143	ttctcaagcctccaaaggaatgatagtcaataaactctgaagcggcgtcttggaatttggaaattca	202
Dh	170	accggagccttcgcaagtactactagaactcttgaggtcttcattggtgcacacctggagttaa	229
Qy	203	ccctatgycgaaggaattatgatacttctccctcttaagaagaccatccagagaggtggtctt	262
Dh	230	ccagatctgagcaacaacggtgacatctctctccggtgcctgtcttgaagaagccacagagct	289
Qy	263	atgtcctcatgtgcctccataacacagctgtagaggaathccctcttggaanaacctgcagatcalca	322
Dh	290	acgtctttagtgcctcttaatacagtttcgttaacctgcgtcccttgaggaaattacgcattatctc	349
Qy	323	ggaggaatctagtactaagaaatctctctatgccttagacagctctatatccactctgtatgcaa	382
Dh	350	gtgggacaacaactttagagagatcgtatagcctctgtgcatactatttttaacctcagaaaaag	409
Qy	383	ataaaac---ggaacgaagagctgcgccaatgagaaatttacagugaaatccctgcctggcg	439
Dh	410	atggaagaacttggacctcttaagaactctgtatthaagaactctgacagaataccccaatctgagc	465
Qy	440	ccgttcggtctcagaacaacaacctgcgtccctctgtcaacagctgtgagagcatccagctgtgcgga	499
Dh	470	ggatctctgtagacagacaagaataatctctcttctgttatgtcagacaacatctatctgtcagaaga	529
Qy	500	tgttcaagcagtgagactctctcaagaaatgtgtatgtagaactctcgaagacacacccctggcgagct	559
Dh	530	tgcttcggaacccaatgcgtctccatctgacactctgtgtctcaacaabaatgtgtatctcaagat	589
Qy	560	gccaagaatgtgatacccaagctgtgcccaatgtagctgcctggtgtctcagagagagaggaact	619
Dh	590	gtgtgagatctgtccataaatactctgt--actgcgcgtgtgtgtgtagagcccaacgaataatcat	646
Qy	620	ggcagaagaatcacaagaataatcatctgtgcgcgaacgaatgtccgggcgcgtctgcgttggaagaat	679
Dh	647	ggcagaactcttgacaagaagcgtgtgtgtgcagagaacaatgtgcagcgagagtgctacagaaactt	706
Qy	680	cccccaagtgatctgtgcacaacaacaaatgtgctgcagcagagctgtgcacagggcccccggagagagcg	739
Dh	707	agctcaagtgagactgcgtgcacatccgaagaatgtgcgtgcgggtgcgtgcgcagaccataaggaacag	766
Qy	740	actgcctgtctgcgcgcaaatctccgaagacgaagccagctgtcagaagatacctgtccccccac	799
Dh	767	actgctcttcgtcgtgaatgaattcaatgaacagtgtagcagctgtctactcagctgtccccaaba	826
Qy	800	tcattgtctcaacaaccacaacacgtaaccaatgtgaatgtgaatacccgagggcaataacagct	859
Dh	827	cccttgtctcaaatcaatcaacacacatcttcaacgtgagcacaattcaatgtcaagaatatacat	886
Qy	860	ctgtgtcccaactgcgtgagaagaatgtctcccccgaattatgtgtgtgtacagagatccacgctgct	919
Dh	887	atgtagaactctctgtctcaagaagaatgtctcaataaactttgtgtgt--agatctcaagtctctc	943
Qy	920	gctctccgaagcctgtgtggtgcgaagaactatgagatgtgagaggaagaacggtctccgaagtata	979
Dh	944	gtctgtcgtgcctgcgtccctgtatctccaabaatctggaatgagaagaataatctggaattatgaat	1003
Qy	980	agagatgtgcgaagaagcgtgtccgcgacaagaatgtgttaacggagatgtgattgtggaattataag	1039
Dh	1004	aaactgtgcactggaactctgtccccaagaagcgttgtgtatgtgcatctggcagaaggtcatttgtgt	1063
Qy	1040	actcactctccataaagtctagatgaattataaacacttcaaaaactgtcactctcatcaatgtg	1099
Dh	1064	caactcagactgtgtgattcttcagataactgtgaacaatcatataaactgtataacagatacaag	1123
Qy	1100	gcgactatccacaatccctgcgcgtgtgagcttttagaggtgtgactctcttcaacatactctctcc	1158
Dh	1124	ggaaattctgactcttctatgaactcgtgtatctatccatctggggagacccttacaatgtccatgtgaagca	1183
Qy	1160	tgtatccagaagaatgtgataatctctgaaaacgtaaaagaataatccaggtctttctgtcta	1219
Dh	1184	ttagaccagagaagaactgtaaagctcttcttcggaacgtacgaagagataaacaaggtcttcctgtgaaca	1243
Qy	1220	ttcaggtctggtcccttaaaaacagagcagactctccatgcctcttggaagaactcgtgaattatatac	1279

QY 500 tagtcaagcagtgactctcagcaaca tgcgacttcagaaacacacttgagcagct 559
 Db 530 tctgtcggaaacccatgagccttccaaacttgactctgtcacaacaatgtagtcaagat 569
 QY 560 gccaaagatgtatgaagcagctcccaactggagcgtgggtgtcagaaagaaagact 619
 Db 590 gtgagcgttgacataagctcgt---acgtgcgtgtgagggagaccacagaataatcatt 646
 QY 620 gccagaactgacaaataatcatctgtgcccagagcagcctccggcgtccgtgagcagt 679
 Db 647 gccagacttgtagaagagcaggtgtgtcagaaacatgtagcagagatgtagcaggaactt 706
 QY 680 cccccagatgactgctgcacaaacacagctgtgtcagagcgtgcagagcccccggagagcgt 739
 Db 707 acgtcagatgactgctgcacaaacacagctgtgtcagagcgtgcagagcccccggagagcgt 766
 QY 740 actgagcgtgtcgtccgcaaatctccagagcagagcagcgtgcagagcccccggagagcgt 799
 Db 767 actgagcgtgtcgtccgcaaatctccagagcagagcagcgtgcagagcccccggagagcgt 826
 QY 800 tcatgtctacaaacccacacacagatgtagatgtgaaacccggaggaataacagct 859
 Db 827 ccttgctctacaaacccacacacagatgtagatgtgaaacccggaggaataacagct 886
 QY 860 ttgtgtccacctgctgtgagaaagatgtcccgtaattatgtgtgtagacagatcagcgtcgt 919
 Db 887 atggaagcattctgtgtcagaaagatgtccacataacttctgtgt---agatcccggtctct 943
 QY 920 ggcgcacagcctgtgtggggccgagctatgagatgtagagagagagcgcgcgcagagctga 979
 Db 944 gtgtgctgtgcgtccctagcttccaaagatgtagaagaaatgtggatlaaagatgtga 1003
 QY 980 agaaaggtgcgaagggcctgtgtgcgaagaaatgtgtgaagaaatgtgtgtgaatlaag 1039
 Db 1004 aactgtgacatgacatctgtgcgaagaaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1063
 QY 1040 acttactctccataaactgtcagaaatlaaacactcaaaacgtcacctccatcagtg 1099
 Db 1064 cagctcagagatgt 1123
 QY 1100 ggcgtctccac 1159
 Db 1124 ggaatttgactcttctagtcacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1183
 QY 1160 tggatccag 1219
 Db 1184 tagagccag 1243
 QY 1220 ttacagctgt 1279
 Db 1244 taccagatgagcacaacaac 1303
 QY 1280 ggcgcagagac 1339
 Db 1304 gtggaagagagac 1363
 QY 1340 ccttgagagatgagcctccacag 1399
 Db 1364 ctctacagcttccagctccacag 1423
 QY 1400 aaaaattgtgtctatgcaaatatcaataaactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1459
 Db 1424 gcaaacctgtgtctatgcaaatatcaataaactgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1483
 QY 1460 aaaaacaaatattatgaacacagagagagagagagagagagagagagagagagagagag 1519
 Db 1484 gaataagatacccggaacacag 1543
 QY 1520 atgactgtgtctcccccag 1579
 Db 1544 accatcgtgtctcccccag 1603
 QY 1580 ggaatgtcagacgagcag 1639

Db 1604 ggcgcctttagtagaagagagatctgacatagagctctcttaactctatagatgtgacttc 1663
 QY 1640 gggagcttctgtgaaactctgagctgacatagagctgacacagagagtg---cctgtccagag 1696
 Db 1664 gggagcttctgtgaaactctgagctgacatagagctgacacagagagtgagagagagagag 1723
 QY 1697 ccatgaaatcactgacag 1756
 Db 1724 atgagcctctcactgacag 1783
 QY 1757 ttgagcgcctcactgagctgacagagagagagagagagagagagagagagagagagagag 1816
 Db 1784 aagatgagcctcactgagctgacagagagagagagagagagagagagagagagagagagag 1843
 QY 1817 tggctggaagatgacag 1876
 Db 1844 ttctc---aagtagctgacatcagatcagagagagagagagagagagagagagagagagag 1900
 QY 1877 acgagtgacacagcggcag 1907
 Db 1901 aaggtgtgaaagcgcacacatgacatgacag 1931

 RESULT 14
 AAA09455
 ID AAA09455 standard; DNA: 3768 BP.
 XX
 AC AAA09455;
 XX
 DT 10-AUG-2000 (first entry)
 XX
 DE Human heregulin 2 (Her2) coding sequence.
 XX
 KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
 KW self-protein; cell-associated peptide antigen; foreign epitope;
 KW cancer; breast cancer; prostate cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WC200020027-A2.
 XX
 PD 13-APR-2000.
 XX
 PF 05-OCT-1999; 99WO-DK00525.
 XX
 PR 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX
 FA (HEBI-) M & E BIOTECH AS.
 XX
 PI Steinae L, Mouritsen S, Nielsen KG, Hanning J, Leach D, Dalum I,
 PI Gautam A, Bltk P, Karlsson G;
 XX
 DR WPI: 2000-349917/30.
 DR P-PSDB: AAY92620.
 PT Inducing immune responses to weakly immunogenic, tumor associated
 XX peptide antigens for the treatment of breast and prostate cancer
 PS Claim 62: Page 187-193; 220pp: English.
 CC The claims detail a method for inducing immune responses against weakly
 CC immunogenic cell-associated peptide antigens (PA) such as those
 CC associated with cancers (i.e., self-proteins), for example, human
 CC prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or
 CC fibroblast growth factor 8b (FGF8b). The method comprises effecting
 CC simultaneous presentation of antigen producing cells (APCs) of the
 CC animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte)
 CC group derived from the PA and/or at least 1 B-cell group derived from the
 CC cell-associated PA; and (2) at least 1 first T helper cell group which is
 CC foreign to the animal. Analogues of human PSM, human Her2 and
 CC human/murine FGF8b comprising a substantial part of all known and

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2002, 19:03:25 ; Search time 1700.56 Seconds
(without alignments)
12852.774 Million cell updates/sec

Title: US-09-715-249-1_COPY_1_2034

Perfect score: 1 atgcagccctcgcgagcgc.....acatcgttcgagagcagc 2034

Sequence: IDENTITY_MUC

Scoring table: Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

1 number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

EST:
1: em_estfun:
2: em_esthum:
3: em_estlin:
4: em_estlom:
5: em_estlpl:
6: em_estlba:
7: em_estlro:
8: em_estloy:
9: em_estlhc:
10: gb_estl:
11: gb_estl2:
12: gb_hlc:
13: gb_gss:
14: em_gss_fun:
15: em_gss_hum:
16: em_gss_iny:
17: em_gss_pln:
18: em_gss_pro:
19: em_gss_rtd:
20: em_gss_vit:
21: em_gss_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1418.4	69.7	2456	12 AK004911	AK004911 Mus muscu
2	1416.8	69.7	2662	12 AK004883	AK004883 Mus muscu
3	1416.8	69.7	2936	12 AK004944	AK004944 Mus muscu
4	1416.8	69.7	2936	12 AK004944	AK004944 Mus muscu
5	739.4	36.4	761	10 A0140518	A0140518 Mus muscu
6	739.4	36.4	761	10 A0140518	A0140518 Mus muscu
7	739.4	36.4	761	10 A0140518	A0140518 Mus muscu
8	620.2	30.5	1035	10 BE742045	BE742045 Mus muscu
9	614.4	30.2	736	10 BE742045	BE742045 Mus muscu
10	611.4	30.1	798	10 BE742045	BE742045 Mus muscu
11	606.8	29.8	634	10 BE740357	BE740357 Mus muscu
12	589	29.0	805	10 A1740805	A1740805 Mus muscu

13	581.2	28.6	757	10 A1935313	A1935313 Mus muscu
14	533.8	26.2	635	10 BE262509	BE262509 Mus muscu
15	523.8	25.8	1031	11 B1332447	B1332447 Mus muscu
16	515	25.3	761	11 B1147172	B1147172 Mus muscu
17	509.4	25.0	678	10 A1554567	A1554567 Mus muscu
18	502.2	24.7	784	10 BE260467	BE260467 Mus muscu
19	459.4	22.6	589	11 BG831511	BG831511 Mus muscu
20	434.4	21.4	582	10 A1858079	A1858079 Mus muscu
21	423.6	20.8	668	11 BE917493	BE917493 Mus muscu
22	423.6	20.8	661	11 BE917493	BE917493 Mus muscu
23	411.8	20.2	1198	11 BE347824	BE347824 Mus muscu
24	399	19.6	1059	11 BG958833	BG958833 Mus muscu
25	395.4	19.4	645	10 A1633375	A1633375 Mus muscu
26	382.8	18.8	415	10 A1763320	A1763320 Mus muscu
27	364	17.9	540	10 BE722865	BE722865 Mus muscu
28	362.4	17.8	759	10 A0139492	A0139492 Mus muscu
29	349.2	17.2	558	10 BE288225	BE288225 Mus muscu
30	347.8	17.1	740	11 B1102081	B1102081 Mus muscu
31	333.2	16.4	464	10 AA701271	AA701271 Mus muscu
32	331.4	16.3	458	11 W14642	W14642 Mus muscu
33	327	16.1	646	10 A1630338	A1630338 Mus muscu
34	324.6	16.0	481	10 A1156625	A1156625 Mus muscu
35	324.4	15.9	446	11 B1330786	B1330786 Mus muscu
36	321	15.8	475	11 BE056131	BE056131 Mus muscu
37	307.8	15.1	955	11 BE533273	BE533273 Mus muscu
38	307	15.1	906	11 BE532802	BE532802 Mus muscu
39	306.2	15.1	906	11 BE532802	BE532802 Mus muscu
40	302.4	14.9	764	10 A1747544	A1747544 Mus muscu
41	299.4	14.7	430	11 BE660321	BE660321 Mus muscu
42	290.6	14.3	502	10 A1263609	A1263609 Mus muscu
43	290	14.1	444	10 BE327106	BE327106 Mus muscu
44	287	14.1	684	10 BE746257	BE746257 Mus muscu
45	287	14.1	935	10 BE745530	BE745530 Mus muscu

ALIGNMENTS

RESULT 1	AK004911	2456 bp	mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300005M11, full insert sequence.	05-JUL-2001
LOCUS	AK004911			
DEFINITION	AK004911			
ACCESSION	AK004911			
VERSION	AK004911.1			
KEYWORDS	cap trap, cDNA library			
SOURCE	Mus musculus			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
AUTHORS	1 (bases 1 to 2456)			
TITLE	Carninci, P. and Hayashizaki, Y.			
JOURNAL	High-efficiency full-length cDNA cloning			
MEDLINE	Methods in enzymology. 303, 19-44 (1995)			
PUBMED	95279253			
REFERENCE	10349636			
AUTHORS	2 (bases 1 to 2456)			
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.			
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
MEDLINE	Genome research. 10 (10), 1617-1630 (2000)			
PUBMED	20499374			
REFERENCE	3 (bases 1 to 2456)			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishikawa, K., Katsunari, T., Tashiro, H., Itoh, M., Saito, N., Ishii, Y., Nakamura, S., Hazama, M., Nishitani, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, H., Fujiwara, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Watanabe, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,			

Db 335 GCAGCTGGGCGGCTTGGAGGAAAGAAAGTCTGCCAAGGCAACATTAACAGCTCACCAG 394
 Oy 121 tctgggactcttgaaagatcatcttctcagcctccagagagatgtaactgaagctg 180
 Db 395 ctgggactcttgaaagatcatcttctcagcctccagagagatgtaactgaagctg 454
 Oy 181 gtccttggaatcttgaaatctcctatctgagagaaatctgaaatctccttctaaag 240
 Db 455 gtccttggaatcttgaaatctcctatctgagagaaatctgaaatctccttctaaag 514
 Oy 241 accatccagagagtgagctgctatctcctatctcctcctcctcctcctcctcctcct 300
 Db 515 accatccagagagtgagctgctatctcctatctcctcctcctcctcctcctcctcct 574
 Oy 301 tctggaaacccgagagatctcctcctcctcctcctcctcctcctcctcctcctcct 360
 Db 575 tctggaaacccgagagatctcctcctcctcctcctcctcctcctcctcctcctcct 634
 Oy 361 gtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 420
 Db 635 atcctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 694
 Oy 421 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 480
 Db 695 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 754
 Oy 481 agcatccagtgagcagagatctcctcctcctcctcctcctcctcctcctcctcctcct 540
 Db 755 actatccagtgagcagagatctcctcctcctcctcctcctcctcctcctcctcctcct 814
 Oy 541 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 600
 Db 815 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 874
 Oy 601 gctgagagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 660
 Db 875 gctgagagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 934
 Oy 661 gctgagagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 720
 Db 935 gctgagagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 994
 Oy 721 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 780
 Db 995 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1054
 Oy 781 aagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 840
 Db 1055 aagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1114
 Oy 841 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 900
 Db 1115 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1174
 Oy 901 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 960
 Db 1175 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1234
 Oy 961 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1020
 Db 1235 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1284
 Oy 1021 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1080
 Db 1295 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1354
 Oy 1081 aagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1140
 Db 1355 aagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1414
 Oy 1141 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1200
 Db 1415 cagagaaatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1474

Oy 1201 atcagagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1260
 Db 1475 atcagagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1534
 Oy 1261 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1320
 Db 1535 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1594
 Oy 1321 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1380
 Db 1595 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1654
 Oy 1381 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1440
 Db 1655 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1714
 Oy 1441 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1500
 Db 1715 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1774
 Oy 1501 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1560
 Db 1775 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1834
 Oy 1561 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1620
 Db 1835 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1884
 Oy 1621 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1680
 Db 1895 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1954
 Oy 1681 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1740
 Db 1955 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 2014
 Oy 1741 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1800
 Db 2015 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 2074
 Oy 1801 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1860
 Db 2075 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 2134
 Oy 1861 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 1920
 Db 2135 gctgagagagagagatctcctcctcctcctcctcctcctcctcctcctcctcct 2182

RESULT 4
 Locus AUI40518 779 bp mRNA EST 25-OCT-2000
 DEFINITION AUI40518 PLACE3 Homo sapiens CDNA clone PLACE300215 5', mRNA
 sequence.
 ACCESSION AUI40518
 VERSION AUI40518.1 GI:11002039
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 779)
 AUTHORS Ota, T., Suzuki, Y., Saito, K., Ishii, S., Yamamoto, J., Sugiyama, T.,
 Nishikawa, T., Nakamura, Y., Sugano, S., Masuno, Y. and Isogai, T.
 HRI human cDNA project (Ota, T., Suzuki, Y., Saito, K., Ishii, S.,
 Yamamoto, J., Sugiyama, T., Nishikawa, T., Nakamura, Y., Sugano, S.,
 Masuno, Y., Isogai, T.)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan


```

OY 1520 atgcttgtgtctccccgagagctctgtgggcccagagccagagctctcttgc 1579
      |||
DB 217 ATGCTTGTGTCTCCCCGAGAGCTCTGTGGGCGGAGGCCA-GGACTGCGTCTTGC 275
      |||
OY 1380 ggaatctcagccgagagcagagatcgctggaagaatgcaagctctcgaagggagccaa 1639
      |||
DB 276 GGAATGTCAAGCCAGGAGGAGATGCGTGCACAAGTCAACCTTCTGAGGCTGAGCCAA 335
      |||
OY 1640 gggagctctggaagaactcagtgcaatagctgcccacacagatgctctcagggca 1699
      |||
DB 336 GGGAGTGTGTGGAGAACTCTGTGATGATACAGTGCACCAAGATGCTCTGCTGAGGCCA 395
      |||
OY 1700 tgaatctcagccgagagcagagagacagacacatgatacagtgctcccaatactg 1759
      |||
DB 396 TGAATCTCACTGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 455
      |||
OY 1760 beggcccccactgctcgaagacctgcccagagatcctggaagaaacacacccctg 1819
      |||
DB 456 ACCGCCCCCACTCCTCAAGACCTGCCGAGAGTCAATGGGAGAAAACAACCCCTG 515
      |||
OY 1820 tctggaagtaagcagagccagccatgtgtgcaaccctgacatccaaactgacactag 1879
      |||
DB 516 TCTGGAAGTACGCAAGACCGCGCATGTGTGCCACTGTGCCATCCAAACTG-ACCTAG 574
      |||
OY 1880 gatgacatgagcagatctctgaagcgtccaaagaaatggagcctaaatcccgatcag 1939
      |||
DB 575 GATGCACTGAGCAGCTCTTGAAGCTGTCCAAAGATGGGCTCAAGATCCCGTCAATG 634
      |||
OY 1940 ccaatggaatgctggggccctcctctctgctgctggtggccctggagatcggcct 1999
      |||
DB 635 CCACTGGAATGCTGGGCGGCGCTCTGCTGTGCTGTGGTGGC-TCGGGATCGCGCT 692
      |||
OY 2000 tcatgcaagagcagcatgctcgaagagcag 2034
      |||
DB 693 TCATGCAAGG-GGCACATCCTTGCAGAGCCACG 726
      |||
RESULT 9
BE315382 736 bp mRNA EST 26-OCT-2000
LOCUS 601141568P1 NIH_MGC_9 Homo sapiens cdna clone IMAGE:3141184 5',
DEFINITION mRNA sequence.
ACCESSION BE315382
VERSION BE315382.1 GI:9146110
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 736)
REFERENCE 1
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rstraus@remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: L10M10 row: h column: 17
High quality sequence stop: 627.
Location/Qualifiers
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3141184"
/clone_1b="NIH_MGC_9"
/issue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally

```

```

BASE COUNT 171 a 200 c 228 g 137 t
ORIGIN
Query Match 30.2% Score 614.4; DB 10; Length 736;
Best Local Similarity 98.8%; Pred.No. 1.7e-149;
Matches 640; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
OY 1369 agtcatggaatgataatctcgaagaacaaatctgtctatgcaaatc-aatnaa 1427
      |||
DB 7 AGTCAATGGAATGTATATTTTCAGGAACAAAATTTCTGTATGCAAAAGCAATATA 66
      |||
OY 1428 ctgaaataacatgcttggaccctcagatcagaataacaaatctaaagcaagatgga 1487
      |||
DB 67 CTGAAATAACATGTTGGACCTCGGTCAAGAAACCAAAATTTATACACAGAGTGA 126
      |||
OY 1488 aacacatgcaagcagcagagccagatgctcagatgctcctcccgagggcagctg 1547
      |||
DB 127 AAGACACTCAAGAGCCACAGGCGCAGGCTCCAGGCTCCCTCCCGAGGCTCTG 186
      |||
OY 1548 gggccggagcccaaggacatgctcctcctgacgaatgtcaagcagaggaatgct 1607
      |||
DB 187 GGGCCGGAGCCCAAGGAGCTCGTCTTTCGGAATGTCAACGAGCAAGCAATGCT 246
      |||
OY 1608 ggaacatgcaagcctcctgagaggtgagcaagagatctgtggaatctgagtgat 1667
      |||
DB 247 GGAACATGCAACCTCTGAGAGGCTGAGCCAAAGGAGTTGTGGAACCTGAGTCA 306
      |||
OY 1668 acaatgcaagcagatgctcctcctcagatgcaatcagcagcagagagagc 1727
      |||
DB 307 ACAGTCCACCCAGAGTCTGCTCTGAGGCAATGAAATCACTGCAACAGAGGAGCC 366
      |||
OY 1728 agacaactglatcagatgtgcccacatctagcagcccccactgctcaagacatgccc 1787
      |||
DB 367 ACACAATGTATTCAGTGTGCGCCACTGATTTGAGGCGCCCACTGCTGCAAGACTGCC 426
      |||
OY 1788 ggaagagatcagtgagagaaacacacacacacacacacacacacacacacacacac 1847
      |||
DB 427 GCGAGAGTCAATGGAGAGAAACACACACCTGCTGGAAGTACGAGACGCCGCTATG 486
      |||
OY 1848 gggccacatgctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1907
      |||
DB 487 GTGCCACCTGTGCTCAATCCAACTGCACTAGGATGCACTGGGCCAGGCTCTGAAAGCTG 546
      |||
OY 1908 tccaagaaatgggctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1967
      |||
DB 547 TCCAAAGATGGGCTTAAATGATCCGCTCAATGCGCACTGGAGATGGTGGGCGCCCTCT 606
      |||
OY 1968 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 2015
      |||
DB 607 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 653
      |||
RESULT 10
A1761585 798 bp mRNA EST 20-DEC-1999
LOCUS A1761585
DEFINITION w66b11.x1 Soares.NSF.FB.9W.OT.PA.P.S1 Homo sapiens cdna clone
IMAGE:2370045 3' similar to gb:K03193 EPRIDRML GROWTH FACTOR
RECEPTOR PRECURSOR (HUMAN);.. mRNA sequence.
ACCESSION A1761585
VERSION A1761585.1 GI:5177252
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 798)

```


(HUMAN); mRNA sequence.

ACCESSION A1935313.1 GI:5674183

VERSION A1935313.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 757)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Unpublished (1997)

Tumor gene index

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.lnl.gov/bio/image/image.html

Insert length: 1220 Std Error: 0.00

Seq primer: -400p from Glibco

High quality sequence stop: 446.

Location/Qualifiers

1..757

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2465052"

/clone_lib="NCI-CGAP_Luig"

/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"

/dev_stage="adult"

/lab_host="DH10B (phage-resistant)"

/note="Organ: Lung; Vector: pRT73-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 177 a 211 c 214 g 154 t 1 others

ORIGIN

Query Match 28.6%; Score 581.2; DB 10; Length 757;
Best local similarity 98.4%; Pred. No. 7.9e-141;
Matches 597; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

1 atcgacccctcgagagcgagcgctctgagcgctctgagcgctctgagcg 60
|||||
2 atcgacccctcgagagcgagcgctctgagcgctctgagcgctctgagcg 60
|||||

Db 146 ATCGACCCCTCGAGAGCGAGCGCTCTGAGCGCTCTGAGCGCTCTGAGCG 205

61 gcgagtcgagcgctcgagagaaagagttgcagagcgagtaacaagctcagcg 120
|||||

Db 206 GCGAGTCGAGCGCTCTGAGAGAAAGAGTTCAGAGCGAGTTCAGAGCGAG 265

121 ttgagacacttgagagcaatcttcagcgctcagagagatgtaacaactgagtg 180
|||||

Db 266 TTGAGACACTTGTGAGAGCAATTTCTCAGCGCTCAGAGAGTTCATTAAC 325

181 gtccctgagagaaattgagaaattgagaaattgagaaattgagaaattgag 240
|||||

Db 326 GTCCCTGAGAGAAATTGAGAAATTGAGAAATTGAGAAATTGAGAAATTGAG 385

241 accatccagagagtgagctgagctcagctcagctcagctcagctcagctc 300
|||||

Db 386 ACCATCCAGAGAGTGAGCTGATGCTCATGCTCATGCTCATGCTCATGCT 445

Qy 301 ttgagc...cctgagagatcagagagaaatgagagaaatcctatgagcga 360

Db 446 TTGGAACACCTGAGAGATCAGAGAAATATGATACGAAATTTCTATGCTTACGA 505

Qy 361 gctcatcactaactgagtgcaaaataaaccgagagagagagctgcccatagaaatla 420

Db 506 GCTTATATCTACTGTGATGTCANNTAAACCGAGCTGAAGAGAGCTGCCATGAGAGAA- 564

Qy 421 cagagacacccctgagctgagcgctgagctgagctgagctgagctgagctgag 480

Db 565 CAGGAATCTCTGCAATGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 624

Qy 481 agcatccagctgagcgagacagacagcagcagcagcagcagcagcagcagcag 540

Db 625 AGCATCCAGCTGAGCGAGATGATGATGATGATGATGATGATGATGATGATGAT 684

Qy 541 gctcagacacccctgagcgagcagcagcagcagcagcagcagcagcagcagcag 600

Db 685 CAGACACACCTGAGCGAGCTGAGCGAGCTGAGCGAGCTGAGCGAGCTGAGCG 744

Qy 601 ggtcgagc 607

Db 745 GGTGCGAG 751

RESULT 14
BE262509 635 bp mRNA EST 26-OCT-2000
LOCUS 601152441F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3508498 5',
DEFINITION mRNA sequence.
ACCESSION BE262509
VERSION BE262509.1 GI:9135588
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 635)
NIH-MGC <http://imgc.ncbi.nlm.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM190 row: a column: 11
High quality sequence start: 12
High quality sequence stop: 631.
Location/Qualifiers

1..635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3508498"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Brain; Vector: pRT73; Site: 1; XhoI; Site: 2; EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH-MGC Library."

BASE COUNT 136 a 187 c 201 g 141 t

ORIGIN

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2002, 20:33:47 ; Search time 2222.52 Seconds
(without alignments)
534.437 Million cell updates/sec

Title: US-09-715-249-1_COPY_1_72
Perfect score: 72
Sequence: 1 atgcgacctccggagcgc.....ctgcgccggcgaagtcgagct 72

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ha:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_pl:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	100.0	560	9	HUMEGFRG M11234 Human epide
2	72	100.0	1347	9	HSEGRF1 X06370 Human A431
3	72	100.0	1593	9	HSU48722 U48722 Human epide
4	72	100.0	1868	6	I80039 Sequence 1
5	72	100.0	1868	9	HSU95089 u95089 Human trunc
6	72	100.0	2643	9	HUMEGFRS K03193 Human abert
7	72	100.0	2864	9	AF125253 Homo sapi
8	72	100.0	3633	6	AX179384 AF125253
9	72	100.0	3888	9	HUMEGFR M38425 Human EGF r
10	72	100.0	3888	11	G31681 SWSS1476 Et
11	72	100.0	5532	6	AR086089 Sequence
12	72	100.0	5532	9	HSRGPRE X00588 Human mRNA
13	72	100.0	17798	9	AC006977 Homo sapi
14	72	100.0	197496	9	AF288738 Homo sapi
15	43.8	60.8	1858	10	AB025197 Rattus no
16	43.8	60.8	1958	10	AF187818 Rattus no
17	43.8	60.8	3208	10	AF142153 Rattus no
18	43.8	60.8	4194	10	AF142153 Rattus norv
19	42.2	58.6	1292	10	RATEGR AF275364S1
20	42.2	58.6	2301	10	MMBGR
21	42.2	58.6	2618	10	AF124513
22	42.2	58.6	5935	10	AF275367 Mus muscu
23	40.6	56.4	3704	10	MMU03425
24	40.6	56.4	4188	10	MMBGR
25	40.6	56.4	163451	2	AC069468 Mus muscu
26	33.6	46.7	1775	9	AF115392
27	33.6	46.7	19527	2	AC067859 Homo sapi
28	32.6	45.3	1280	6	ARI28142 Sequence
29	32.6	45.3	2648	5	ARI28142 Chicken hsp
30	32.6	45.3	16188	1	AE006968 Mycobacte
31	32.6	45.3	28826	1	AE006968 Mycobacte
32	31.8	44.2	2223	1	MXA133131 Myxococu
33	31.4	43.6	35654	1	SC7H1
34	30.6	42.5	5719	2	AC021328
35	30.6	42.5	5719	2	AF327569 Mus muscu
36	30.6	42.5	204412	2	AC021446 Mus muscu
37	30.4	42.2	18161	2	AC073936 Mus muscu
38	30.2	41.9	143894	33	AC069002
39	30.2	41.9	156148	2	AL591038
40	30.2	41.9	182486	2	AC073209
41	30	41.7	5210	9	AB023137 Homo sapi
42	30	41.7	7525	9	AJ303079 Homo sapi
43	30	41.7	83001	2	AC015632 Homo sapi
44	30	41.7	89813	2	AC023660 Homo sapi
45	30	41.7	110000	2	AL359978 Homo sapi

ALIGNMENTS

RESULT	1	PRI	07-NOV-1994
LOCUS	HUMEGFRG	560 bp	DNA
DEFINITION	Human epidermal growth factor receptor (EGFR) gene, exon 1.		
ACCESSION	M11234		
VERSION	M11234.1 GI:181981		
KEYWORDS	epidermal growth factor receptor.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 560)		
AUTHORS	Ishii,S., Xu,Y.H., Stratton,R.H., Roe,B.A., Merlino,G.T. and Pastan,I.		
TITLE	Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 82 (15), 4920-4924 (1985)		
MEDLINE	85270438		
FEATURES	Location/Qualifiers		
SOURCE	1..560		

	Query Match	Score	72:	DB	97:	Length	560;
	Best Local Similarity	100.0%	Pred.	No.	4.1e-07:		
Matches	72:	Conservative	0:	Mismatches	0:	Indels	0:
OY	1 atggagacccctcaggagaccgagcaaacgccttcgtcgatgatgaacttcccc	60					
Dd	466 ATGCGACCCCTCCGGGAGCGGCCGGGCAGCGCCTCTGTGGCCTTGCTTCCCC	525					
OY	61 gtagagtctgggct	72					
Dd	526 GCGAGTCGGGCT	537					

mRNA	/number=1	839..>1189
precursor_RNA		839..>1347
precursor_RNA	/note="primary transcript"	839..>1347
misc_feature	/note="altern. primary transcript"	925
precursor_RNA	/note="5' end of longest cDNA clone (see X00588)"	928..>1347
precursor_RNA	/note="altern. primary transcript"	939..>1347
precursor_RNA	/note="altern. primary transcript"	979..>1347
precursor_RNA	/note="altern. primary transcript"	992..>1347
precursor_RNA	/note="altern. primary transcript"	999..>1347
CDS	/note="altern. primary transcript"	1102..>1189
	/note="A431 EGF receptor (AA 1 - 29) (1189 is 1st base in codon)"	
	/codon_start=1	
	/protein_id="CAA29668.1"	
	/db_xref="GI:31119"	
	/db_xref="SWISS-PROT:P00533"	
	/translation="MRPSGTAGALLALALCAPASRALEEK"	
	/number=1	1190..>1347
intron		1216..1221
misc_feature	/note="pot. Spl binding site"	1248..1251
misc_feature	/note="pot. Spl binding site"	1285..1291
misc_feature	/note="pot. Spl binding site"	1398 g 231 t
BASE COUNT	210 a 508 c 398 g 231 t	
ORIGIN		
Query Match	100.0%:	Score 72: DB 9: Length 1347:
Best Local Similarity	100.0%:	Pred. NO. 3e-07:
Matches 72: Conservative	0: Mismatches	0: Indels 0: Gaps 0:
Oy 1 atgcgacccctcggagcgaggcgagcgcgtcgtgagctgctgagctgcgccg 60		
Dd 1102 ATGCGACCCCTCCGGAGCGGCCGGCAGCCGCCTCGGCGCTGCGCTGCCCCG 1161		
Oy 61 gcgagtcggact 72		
Dd 1162 GCGAGTCGGGCT 1173		
RESULT 3		
LOCUS HSU48722	1593 bp mRNA PRI	05-NOV-1996
DEFINITION Human epidermal growth factor receptor precursor (EGFR) mRNA,		
complete cds.		
VERSION U48722		
KEYWORDS U48722.1 GI:1628549		
SOURCE human.		
ORGANISM Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
REFERENCE 1 (bases 1 to 1593)		
AUTHORS Reiter,J.L. and Mainle,N.J.		
TITLE A 1.8 kb alternative transcript from the human epidermal growth		
JOURNAL factor receptor gene encodes a truncated form of the receptor		
MEDLINE Nucleic Acids Res. 24 (20), 4050-4056 (1996)		
REFERENCE 2 (bases 1 to 1593)		
AUTHORS Reiter,J.L.		
TITLE Direct Submission		
JOURNAL Submitted (07-FEB-1996) Jili L. Reiter, Biochemistry and Molecular		

FEATURES Biology, Mayo Clinic, 200 First St. SW, Rochester, MN 55905, USA
location/Qualifiers
1 1003

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p11.2-p12"
/tissue_type="placenta"
1..1456
/gene="ECFR"
1..244
/gene="BGRF"
245..1462
CDS

```

```

/product="epidermal growth factor receptor precursor
/protein_id="AAC50802.1"
/db_xref="GI:1639850"

```

/krienslatlan-n- mfrsragagallailaalaacspkralaeekvcoqsnitltoigfr
 edheslajorm-nen-evualjelttyoyrnydailcftltoevygalvialnteltoigfr
 nlolirgmyayensyalavysindankrtdelkelpmnlnoelmgavnsrnpalncavne
 siowrd iysafssylsnm5delohng5c5qcd5pc5oc5lga5vrc5or5lk5it5caod5
 c5gr5c5r5p5cd5hnc5aag5c5p5d5c5oc5lga5vrc5or5lk5it5caod5
 m5vne5c5y5sga5c5cr5rny5v5p5d5c5oc5lga5vrc5or5lk5it5caod5
 k5vnc5it5g5r5c5p5d5hnc5aag5c5p5d5c5oc5lga5vrc5or5lk5it5caod5
 k5vnc5it5g5r5c5p5d5hnc5aag5c5p5d5c5oc5lga5vrc5or5lk5it5caod5

```

sig_peptide      245..316
                  /gene="EGFR"
mat_peptide      317..1459
                  /product="epidermal growth factor receptor"
misc_feature     626..634
                  /gene="EGFR"

```

misc_feature	misc_feature	misc_feature	misc_feature
767..775	767..775	767..775	767..775
/note="encodes Asn-linked glycosylation site"	/note="encodes Asn-linked glycosylation site"	/note="encodes Asn-linked glycosylation site"	/note="encodes Asn-linked glycosylation site"
/gene="EGFR"	/gene="EGFR"	/gene="EGFR"	/gene="EGFR"
830..838	830..838	830..838	830..838

misc_feature	/note="encodes Asn-Linked glycosylation site"
misc_feature	1055..1063
	/gene="EGFR"
misc_feature	/note="encodes Asn-Linked glycosylation site"
	1298..1306

```
misc_feature 1325..1333
              /note="encodes Asn-linked glycosylation site"
              /gene="EGFR"
misc_feature 1452..1456
              /note="encodes Asn-linked glycosylation site"
```

	3' UTR	polya.signal	polya.site	BASE COUNT	ORIGIN
1				406 a	445 c
2					424 g
3					318 t

Query Match	100.0%;	Score 72;	DB 9;	Length 1593;
Best Local Similarity	100.0%;	Pred. No. 2.8e-07;		
Matches 72;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Oy	1	atcggacccctccggagacgagccggcgagcgcgtctccctgacgtctgtgtgctcgcgtctcgcgcg	60
Db	245	ATGGACCCCTCCGGAGACGGCGGGGACACGCTCCTGGCGCTGCTGGCTGCGCTTGCCCG	304
Oy			
Oy	61	ggcagatcggact	72
Db	305	GGCAGATGGACT	316

RESULT 4
I80039

LOCUS	80039	1868 bp	DNA	PAT	10-JUN-1998
DEFINITION	Sequence 1 from patent US 5708156.				
ACCESSION	180039				
VERSION	180039.1	GI:3208329			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
	Unclassified.				
REFERENCE	1 (bases 1 to 1868)				
AUTHORS	Ilekts,J.V.				
TITLE	Epidermal growth factor receptor-like gene product and its uses				
JOURNAL	Patent: US 5708156-A 1 13-JAN-1998;				
FEATURES	Location/Qualifiers				
source	1..1868				
	/organism="unknown"				
BASE COUNT	449 a	554 c	494 g	371 t	
ORIGIN					

	Query Match	100.0%;	Score 72;	DB 6;	Length 1866;	
	Best Local Similarity	100.0%;	Pred. 2.7e-07;			
	Matches	72;	Conservative	0;	Mismatches	0; Indels
						Gaps
Oy	1 atagacaccctccgcgggagcgcgcgtcctcgagcgtcgtcgtctatgccg	60				
Dd	504 ATGCGACCCCTCGGGAGCGGCCGCAAGCGCTCTGCGCCTGTGGCTTGCCTCTGCCCG	563				
Oy	61 gtagtcgagtcgagtc	72				
Dd	564 GCGAGTCGAGGCT	575				

RESULT	5
HSU95089	
LOCUS	HSU95089 1868 bp mRNA
DEFINITION	Human truncated epidermal growth factor receptor-like protein
ACCESSION	U95089
VERSION	U95089.1 GI:2051984
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 1868)
AUTHORS	Ilekis,J.V., Stark,B.C. and Scoccia,B.
TITLE	Possible role of variant RNA transcripts in the regulation of epidermal growth factor receptor expression in human placenta
JOURNAL	Mol. Reprod. Dev. 41 (2), 149-156 (1995)
MEDLINE	95382957
REFERENCE	2 (bases 1 to 1868)
AUTHORS	Ilekis,J.V., Garitt,J., Niederberger C. and Scoccia,B.
TITLE	Expression of a Truncated Epidermal Growth Factor Receptor-Like Protein (TEGFR) in Ovarian Cancer
JOURNAL	Gyn. Onc. 65 (1997) In press
REFERENCE	3 (bases 1 to 1868)
AUTHORS	Ilekis,J.V.
TITLE	Direct Submission
JOURNAL	Submitted (25-MAR-1997) Obstetrics & Gynecology, University of Illinois at Chicago, 820 S. Wood Street, Chicago, IL 60612, USA
COMMENT	please also see Ilekis, J. in RK Miller, HA Thiede (eds): 'Molecular Biology and Cell Regulation of the Placenta.' New York: Plenum, 103-117, 1991.
FEATURES	
SOURCE	1. 1868 Location/Qualifiers

sig_peptide
CDS

/note="TEGERF: the calculated molecular weight of the nascent protein is approximately 42 kD (includes signal peptide at amino acid positions 1-24); the molecular

ORIGIN Map position 7q12-14.

Query Match 100.0%; Score 72; DB 9; Length 3888;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcgaccctccgagagccgagcagcagcctccgctgctgctgctgctgcccg 60
|||||
Db 1114 ATGCGACCTCCGCGGAGCGCGGCGCGCTCTGCGCGCTGCGCTGCGCGCG 1173
QY 61 gcgagtcggcct 72
|||||
Db 1174 GCGAGTCGGGCT 1185

RESULT 10

LOCUS G31681 3888 bp DNA STS 28-SEP-1998
DEFINITION SMS1476 Eric D. Green Homo sapiens STS genomic, sequence tagged
site.
ACCESSION G31681
VERSION G31681.1 GI:1916406
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3888)
Bouffard G.G., Iyer, L.M., Idol, J.R., Braden, V.V., Cunningham, A.F.,
Weintrub, L.A., Mohr-Tidwell, R.M., Peluso, D.C., Fulton, R.S.,
Lectie, M.P. and Green, E.D.
A collection of 1814 human chromosome 7-specific STS
Genome Res. 7 (1), 59-64 (1997)

TITLE 2 (bases 1 to 3888)
MEDLINE Green, E.D.
REFERENCE Human chromosome 7 STS (1997)
AUTHORS Unpublished (1997)
JOURNAL
TITLE
JOURNAL
COMMENT

Contact: Eric D. Green
Genome Technology Branch
National Human Genome Research Institute/NIH
49 Convent Dr., MSC431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892
Tel: 3014020201
Fax: 3014024735
Email: egreen@nhgri.nih.gov
Primer A: GTTCAAGATTGAGCCACAC
Primer B: TTCTCTGCACACTTGGCAC
STS size: 119
PCR profile:

Presoak: 0 degrees C for 0.00 minute(s)
Denaturation: 92 degrees C for 0.17 minute(s)
Annealing: 60 degrees C for 1.00 minute(s)
Polymerization: 72 degrees C for 1.00 minute(s)
PCR Cycles: 35
Thermal Cycler: PerkinElmer 9600

Protocol:
Template: 30-100 ng
Primer: each 1 uM
dNTPs: each 200 uM
Tag Polymerase: 0.05 units/uI
Total Vol: 10 uI

Buffer:
MgCl2: 1.5 mM
KCl: 100 mM
Tris-HCl: 10 mM
NH4Cl: 5 mM
pH: 8.6

This STS has been incorporated into the NHGRI chromosome 7
physical map, but was developed by another investigator. See
GenBank record: M38425 For additional information about the NHGRI
chromosome 7 mapping project, see
http://www.nhgri.nih.gov/DIR/STH/CHR7. Also see Genomics
11:548-64 (1991) [MUID-92128937].
Location/Qualifiers

FEATURES
source
1..3888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7"
/clone_lib="Eric D. Green"
1..3888
gene
/gene="EGFR"
2358..2456
/gene="EGFR"
2358..2359
/gene="EGFR"
complement(2437..2456)
primer_bind
1219 c 1101 g 806 t
BASE COUNT 762 a 1219 c 1101 g 806 t
ORIGIN

Query Match 100.0%; Score 72; DB 11; Length 3888;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcgaccctccgagagccgagcagcagcctccgctgctgctgctgctgcccg 60
|||||
Db 1114 ATGCGACCTCCGCGGAGCGCGGCGCGCTCTGCGCGCTGCGCTGCGCGCG 1173
QY 61 gcgagtcggcct 72
|||||
Db 1174 GCGAGTCGGGCT 1185

RESULT 11
LOCUS AR086089 5532 bp DNA PAT 07-SEP-2000
DEFINITION Sequence 3 from patent US 5985553.
ACCESSION AR086089
VERSION AR086089.1 GI:10012855
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5532)
AUTHORS Kling, C. Richter, Kraus, M.H. and Aaronson, S.A.
TITLE erbB-2 gene segments, probes, recombinant DNA and kits for
detection
JOURNAL Patent: US 5985553-A 3 16-NOV-1993;
location/Qualifiers

FEATURES
source
1..5532
/organism="unknown"
BASE COUNT 1472 a 1484 c 1337 g 1239 t
ORIGIN

Query Match 100.0%; Score 72; DB 6; Length 5532;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atcgaccctccgagagccgagcagcagcctccgctgctgctgctgctgcccg 60
|||||
Db 187 ATGCGACCTCCGCGGAGCGCGGCGCGCTCTGCGCGCTGCGCTGCGCGCG 246
QY 61 gcgagtcggcct 72
|||||
Db 247 GCGAGTCGGGCT 258

RESULT 12
HSDCFRE

REFERENCE 5 (bases 1 to 177998)

AUTHORS

Waterston, R.

TITLE

Direct Submission

COMMENT

Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Sep 28, 1999 this sequence version replaced gi:4662682.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: saplens@wustl.wustl.edu

----- Summary Statistics

Center project name: R_DJ1091E12

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pterer de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pterer de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone RP5-1091E12 from base position 192688 to 192794 is an approximate representation of a tandem repeat. There are approximately 800 bases of this tandem missing according to the restriction digests: ecorRV band size real at 5047, insillico at 4288; hindIII band size real at 10980, insillico at 10434; bamHI band size real at 5010, insillico at 4180.

The clone sequenced to the left is RP4-791C19, 200 bp overlap. Actual start of this clone is at base position 80354 of RP4-791C19; actual end is at base position 177998 of RP5-1091E12.

FEATURES

source

1..177998

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7p11.2-p21"

/clone="RP5-1091E12"

/clone_id="RPCI-5"

1..2564

/rpt_family="L1"

repeat_region

2565..2617

/rpt_family="Alu"

repeat_region

2744..3029

/rpt_family="Alu"

```

repeat_region 3037..3132
/rpt_family="L1"
repeat_region 3138..3497
/rpt_family="MAlR"
repeat_region 3586..4083
/rpt_family="L1"
repeat_region 4084..4302
/rpt_family="Alu"
repeat_region 4303..4398
/rpt_family="L1"
repeat_region 4417..4476
/rpt_family="MAlR"
repeat_region 4729..5355
/rpt_family="L1"
misc_feature 8556..10519
/clone="CpG island (MGC=72.4, o/e=0.90, #CpGs=225)"
repeat_region 9010..9102
/rpt_family="CT-rich"
misc_feature 9109..9472
/clone="similar to Mus musculus EST A1787580 (NID:g5335296)
u130a11.x1"
repeat_region 9225..9246
/rpt_family="GC-rich"
misc_feature 9231..9472
/clone="match to EST A1554567 (NID:g4486930) t25f08.x1"
misc_feature 9231..9472
/clone="match to EST A1740805 (NID:g5109093) w924d08.x1"
misc_feature 9238..9472
/clone="match to EST AA934507 (NID:g3091664) o050d06.s1"
misc_feature 9240..9472
/clone="match to EST A1935313 (NID:g5674183) w16g07.x1"
misc_feature 9258..9472
/clone="match to EST A1858079 (NID:g5511695) wj70b02.x1"
misc_feature 9261..9477
/clone="similar to EST A1263609 (NID:g3871812) qg89g02.x1"
misc_feature 9268..9472
/clone="similar to EST R35665 (NID:g792366) yH90f01.s1"
repeat_region 9487..9582
/rpt_family="GC-rich"
misc_feature 10633..11006
/clone="match to EST R35059 (NID:g791960) yH86d10.r1"
repeat_region 10667..10700
/rpt_family="CA)n"
misc_feature 10668..10845
/clone="match to EST AA768080 (NID:g2819095) nV96h04.s1"
repeat_region 10949..11138
/rpt_family="MERL-type"
misc_feature 11173..11567
/clone="match to EST R77147 (NID:g851779) y164h11.s1"
repeat_region 11895..12057
/rpt_family="MTR"
repeat_region 13289..13321
/rpt_family="(TG)n"
repeat_region 13475..13776
/rpt_family="Alu"
repeat_region 15231..15735
/rpt_family="L2"
repeat_region 16116..16266
/rpt_family="MERL-type"
repeat_region 16517..16973
/rpt_family="L1R19A"
repeat_region 17021..17182
/rpt_family="MERL-type"
misc_feature 17638..18082
/clone="match to EST A1870800 (NID:g5544768) w169a08.x1"
misc_feature 17959..18068
/clone="match to EST AA640700 (NID:g5565950) n122f01.r1"
repeat_region 18099..18150
/rpt_family="(TA)n"
repeat_region 18838..18872
/rpt_family="AT-rich"
repeat_region 18985..19279
/rpt_family="Alu"

```

[illegible]

REFERENCE	TITLE	JOURNAL	MEDLINE	REFERENCE	FEATURES
PUBMED 97078696					source
4 (bases 192735 to 195366)					
Eley,G., Frederick,L., Wang,X.Y., Smith,D.I. and James,C.D.					
3' end structure and rearrangements of EGFR in glioblastomas					
Genes Chromosomes Cancer 23 (3), 248-254 (1998)					
99005167					
PUBMED 9796506					
5 (bases 1 to 197496)					
Reiter,J.L., Threadgill,D.W., Eley,G.D., Strunk,K.E.,					
Danielsen,A.J., Schell, Sinclair,C., Pearisall,R.S., Green,P.J.,					
Yee,D., Lamond,A.L., Balasubramaniam,S., Crossley,T.D.,					
Magnuson,T.R., James,C.D. and Melhrie,N.J.					
Comparative genomic sequence analysis and isolation of human and					
mouse alternative egr transcripts encoding truncated receptor					
isoforms 71 (1), 1-20 (2001)					
2100872					
6 (bases 1 to 197496)					
Reiter,J.L. and Eley,G.D.					
Direct Submission					
Submitted (21-JUL-2000) Biochemistry and Molecular Biology, Mayo					
Clinic, 200 First St. SW, Rochester, MN 55905, USA					
Part of this sequence is also presented in Genbank Accession number					
AC006977, produced as a collaborative effort between the Genome					
Sequencing Center at the Washington University School of Medicine					
and J.L. Reiter and G.D. Eley.					
location/Qualifiers					
1..197496					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/chromosome="7"					
/map="7p11.2"					
complement(1..2564)					
/rpt_family="L1"					
/rpt_type-dispersed					
complement(2566..2617)					
/rpt_family="Alu"					
2744..3029					
/rpt_type-dispersed					
2744..3029					
/rpt_family="Alu"					
/rpt_type-dispersed					
3037..3132					
/rpt_family="L1"					
/rpt_type-dispersed					
3138..3497					
/rpt_family="MT2A1"					
/rpt_type-dispersed					
complement(3586..4083)					
/rpt_family="L1"					
/rpt_type-dispersed					
4084..4302					
/rpt_family="Alu"					
/rpt_type-dispersed					
complement(4303..4398)					
/rpt_family="L1"					
/rpt_type-dispersed					
complement(4417..4726)					
/rpt_family="MSND"					
/rpt_type-dispersed					
complement(4729..5555)					
/rpt_family="L1"					
/rpt_type-dispersed					
6512..10515					
/note="CpG Island"					
8627..12178					
/note="MSST17c"					
/db_xref="dbSTS:31681"					
194919219..8472..323293					
132544..133412					
141469..142653					
143771..144766					
144118..144259					
145357..146083					
146040..146945					
146866..146939					
147770..150445					
151606..151738					
153840..153930					
153587..153544					

[illegible]

FEATURES	source
gene	1858
CDS	1775..1858
	/gene="egfr"
	/codon_start=1
	/product="egfr"
	/protein_id="BA076391.1"
	/db_xref="GI:4586456"
	/translation="MRPSGATARRKLLLLAALCAAGALEEK"
	424 a 510 c 471 g 453 e

•
•
•
•

•

• •

PA (UYJO) UNIV JOHNS HOPKINS.
 XX Bligner D, Vogelstein B;
 XX WPI: 2000-601693/57.
 DR P-PSDB: AAB19259.
 XX
 PT Diagnosis of glioma in brain cell sample involves determining the
 PT presence of a nucleic acid encoding an epidermal growth factor receptor
 PT mutant protein type II -
 XX
 PS Disclosure: Fig 11A-F; 51pp; English.
 XX
 CC The specification describes a method for diagnosing glioma in a brain
 CC cell sample suspected of being cancerous. The method comprises detecting
 CC the presence of a nucleic acid encoding an epidermal growth factor
 CC receptor (EGFR) mutant protein type II. Deletions in the EGFR gene
 CC are found in many gliomas, breast tumours and lung tumours. The method
 CC is useful for diagnosing gliomas, breast tumours and lung tumours. The
 CC present sequence encodes a normal EGFR polypeptide.

Sequence 3818 BP; 972 A; 1089 C; 1012 G; 745 T; 0 other;

Query Match 100.0%; Score 72; DB 21; Length 3818;

Best Local Similarity 100.0%; Pred. No. 7.1e-10;

Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgcgacccctccgcgagcgccgagcgctcctgcgctgcgtgcgtgcgcgcg 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 187 atgcgacccctccgcgagcgccgagcgctcctgcgctgcgtgcgtgcgcgcg 246

OY 61 gcgagtcgggct 72
 ||||||||||||
 Db 247 gcgagtcgggct 258

RESULT 6
 AA23954
 ID AA23954 standard; DNA: 5532 BP.
 XX
 AC AA23954;
 XX
 DT 28-JAN-2000 (first entry)
 XX
 DE Human EGF receptor DNA.
 XX
 DE Human: v-erbB; MHC117; epidermal growth factor receptor; anticancer;
 XX cytostatic; toxin; cancer; treatment; detection; mammary carcinoma;
 XX malignant; EGF receptor; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FT Key Location/Qualifiers
 FT 187..3819
 FT CDS /*tag= a
 FT /*product= "EGF receptor"

US5985553-A.
 PD 16-NOV-1999.
 XX
 PD 07-JUN-1995; 95US-0475035.
 XX
 PR 21-OCT-1987; 87US-0110791.
 PR 01-NOV-1991; 91US-0786598.
 PR 05-MAR-1986; 86US-0836414.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Aaronson SA, Kraus MH, King CR.
 PI WPI: 2000-012777/01.
 XX
 DR

DR P-PSDB: AAY50616.
 XX
 PY Purified nucleic acid useful for detection and treatment of mammary
 PT carcinoma -
 XX
 PS Disclosure: Column 19-30; 32pp; English.
 XX

CC This invention describes a novel purified nucleic acid (I) (v-erbB
 CC related gene) specifically hybridizing to all or part of a MHC117 gene
 CC and not hybridizing to nucleic acid encoding an epidermal growth factor
 CC receptor. The product of the invention has anticancer and cytostatic
 CC activity. Antibodies to the protein encoded by (I) are conjugated to
 CC toxins and kill cancer cells expressing (I). Antibodies to the protein
 CC encoded by (I) are useful for the treatment of cancer. Fragments of (I)
 CC and the MHC117 gene are useful as probes for the detection of human
 CC mammary carcinoma or other malignancies resulting from the v-erbB related
 CC gene. This sequence encodes a human epidermal growth factor (EGF)
 CC receptor described in the invention.
 CC
 SQ

Sequence 5532 BP; 1472 A; 1484 C; 1337 G; 1239 T; 0 other;

Query Match 100.0%; Score 72; DB 21; Length 5532;

Best Local Similarity 100.0%; Pred. No. 7e-10;

Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgcgacccctccgcgagcgccgagcgctcctgcgctgcgtgcgtgcgcgcg 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 187 atgcgacccctccgcgagcgccgagcgctcctgcgctgcgtgcgtgcgcgcg 246

OY 61 gcgagtcgggct 72
 ||||||||||||
 Db 247 gcgagtcgggct 258

RESULT 7
 AAQ43814
 ID AAQ43814 standard; DNA: 5750 BP.
 XX
 AC AAQ43814;
 XX
 DT 20-OCT-1993 (first entry)
 XX
 DE pRLD3D4 construct.
 XX
 DE Epidermal growth factor receptor truncate protein; EGF;
 KW binding sites; adsorptive agents; mammalian cell growth abnormality;
 KW detection; growth; reproduction; signal transmission; ds.
 XX
 OS Synthetic.
 XX
 XX
 FT Key Location/Qualifiers
 FT misc_feature 1077..1079
 FT /*tag= a
 FT /*note= "codon ARA encodes Ile"

US5218090-A.
 PD 08-JUN-1993.
 XX
 PD 12-JUN-1990; 90US-0536896.
 XX
 PR 12-JUN-1990; 90US-0536896.
 PR 26-OCT-1990; 90US-0604728.
 XX
 PA (WARN) WARNER LAMBERT CO.
 PA
 PI Connors RW;
 PI WPI: 1993-196297/24.
 DR P-PSDB: AAR38211.
 XX
 PT New epidermal growth factor receptor truncate proteins - which

FT	/*tag= d	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 27; encodes AAB07559"	
FT	5806..12294	
FT	/*tag= e	
FT	/note= "ORF 26; encodes AAB07560"	
FT	12291..15491	
FT	/*tag= f	
FT	/note= "ORF 25; encodes AAB07561"	
FT	15488..21013	
FT	/*tag= g	
FT	/note= "ORF 24; encodes AAB07562"	
FT	21010..24666	
FT	/*tag= h	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 23; encodes AAB07563"	
FT	24663..32690	
FT	/*tag= i	
FT	/note= "ORF 22; encodes AAB07564"	
FT	32893..34830	
FT	/*tag= j	
FT	/note= "ORF 21; encodes AAB07565"	
FT	34827..35804	
FT	/*tag= k	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 20; encodes AAB07566"	
FT	35818..37302	
FT	/*tag= l	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 19; encodes AAB07567"	
FT	37299..39215	
FT	/*tag= m	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 18; encodes AAB07568"	
FT	39301..47181	
FT	/*tag= n	
FT	/note= "ORF 17; encodes AAB07569"	
FT	47178..49985	
FT	/*tag= o	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 16; encodes AAB07570"	
FT	49982..51001	
FT	/*tag= p	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 15; encodes AAB07571"	
FT	50998..52386	
FT	/*tag= q	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 14; encodes AAB07572"	
FT	52383..52946	
FT	/*tag= r	
FT	/note= "ORF 13; encodes AAB07573"	
FT	53018..54190	
FT	/*tag= s	
FT	/note= "ORF 12; encodes AAB07574"	
FT	54187..55824	
FT	/*tag= t	
FT	/note= "ORF 11; encodes AAB07575"	
FT	55821..56093	
FT	/*tag= u	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 10; encodes AAB07576"	
FT	56090..57586	
FT	/*tag= v	
FT	/transl_except= (pos: 1..3, aa: Met)	
FT	/note= "ORF 9; encodes AAB07577"	
FT	57583..58857	
FT	/*tag= w	
FT	/note= "ORF 8; encodes AAB07578"	
PN	WO200040704-A1.	
XX		
PD	13-JUL-2000.	

PF	XX	06-JAN-2000;	2006WO-US00445.			
XX	XX	06-JAN-1999;	99US-0115435.			
PR	XX	05-FEB-1999;	99US-0118848.			
PR	XX	05-JAN-2000;	2000US-0477962.			
XX	XX					
PA	XX	(REGC) UNIV CALIFORNIA.				
PI	XX	Shen B, Du L, Sanchez C, Chen M, Edwards DJ;				
DR	XX	WPI: 2000-465974/40.				
DR	XX	P-PSB: AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,				
DR	XX	AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,				
DR	XX	AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,				
DR	XX	AAB07576, AAB07577, AAB07578.				
PT	XX	New bleomycin gene cluster components useful for peptide and/or				
PT	XX	polyketide metabolites, especially bleomycin, production and for				
PT	XX	chemically modifying biological molecules -				
XX	XX	Claim 8: Page 97-136; 162pp: English.				
PS	XX					
CC	XX	The present sequence represents the BLM (bleomycin) gene cluster,				
CC	XX	containing open reading frames (ORFs) 8-30. The proteins encoded				
CC	XX	by the gene cluster are useful for producing peptides and/or polyketide				
CC	XX	metabolites, especially bleomycin or bleomycin analogues. They are				
CC	XX	also useful for chemically modifying biological molecules to produce				
CC	XX	branched methyl groups, and for coupling amino acids and fatty				
CC	XX	acids. They may be reacted with an apo-carrier protein and coenzyme A				
CC	XX	to produce a halo-carrier protein. The BLM gene cluster or catalytic				
CC	XX	domains can be used individually or collectively to produce				
CC	XX	thiazolidine, thiazoline, bithiazoline and bithiazoline-containing				
CC	XX	microbial metabolites. The BLM gene cluster may also be used to produce				
CC	XX	sugars.				
SO	XX	Sequence 58857 BP: 7256 A; 25139 C; 19353 G; 7109 T; 0 other;				
Query Match		40.6%; Score 29.2; DB 21; Length 58857;				
Best Local Similarity		69.0%; Pred. NO. 25;				
Matches	40;	Conservative	0; Mismatches	18; Indels	0; Gaps	0
QY	14	ggacggcggggagcgcgcctctcgtgcgcgtcgtcgtcgtcgtccgcggagtcgggc	71			
DB	10786	GGACGGCGCGCGCCGCCGCTCCCGCGCATGACGCGCGCGAGCGCGCTCGTGGGCTGGGC	10729			
RESULT	12					
ID	AAV23486/C					
XX	AAV23486 standard; DNA; 390 BP.					
AC	AAV23486;					
XX						
DT	24-JUL-1998 (first entry)					
XX						
DE	Pseudomonas XcpV secretion factor coding sequence.					
XX						
KW	Kinase: Lipo; LipR; lipase expression regulator; DNA binding regulator;					
KW	sigma 54 promoter; secretion factor; lux-box binding element;					
KW	oriV-box binding element; regulation cascade; ss.					
XX						
OS	Pseudomonas alcaligenes.					
XX						
PN	WO9806836-A2.					
XX						
PD	19-FEB-1998.					
XX						
PF	15-AUG-1997; 97WO-US14450.					
XX						
PR	16-AUG-1996; 96US-0699092.					
XX						
PA	(GENV) GENENCOR INT' INC.					


```

XX PI Gerritse G, Quax WJ;
XX XX
DR WP1: 1998-159528/14.
XX P-PSDB; AAM53834.
XX XX
PT Nucleic acids encoding proteins involved in the lipase regulation
PT cascade from P. alcaligenes - useful for controlling production and
PT secretion of heterologous proteins in P. alcaligenes
XX PS Claim 13; Page 37; 106pp; English.
XX XX
CC This sequence encodes the XcpV secretion factor of Pseudomonas
CC alcaligenes. The DNA represents a nucleic acid of the invention. The
CC nucleic acids encode: (a) a kinase from a Pseudomonad that regulates the
CC expression of a lipase; (b) a DNA binding regulator from a Pseudomonad
CC that regulates the expression of a lipase; (c) a Pseudomonas alcaligenes
CC upstream activating sequence; (d) a P. alcaligenes sigma 54 promoter that
CC regulates the expression of a lipase; (e) a P. alcaligenes secretion
CC factor selected from XcpP, O, R, S, T, U, V, W, X, Y and Z, and OrfV, X,
CC Y, (f) a P. alcaligenes lux-box binding element; and (g) a orfV-box
CC binding element. The nucleic acids represent parts of a regulation
CC cascade, comprising at its heart a kinase and a DNA binding regulator.
CC These sequences can be used for the production of heterologous proteins
CC in a host cell. The cascade also comprises secretion factors which can
CC enhance the secretion of produced proteins.
XX SQ Sequence 390 BP; 57 A; 127 C; 154 G; 52 T; 0 other:

Query Match          39.7%; Score 28.6; DB 19; Length 390;
Best Local Similarity 67.8%; Pred. No. 44;
Matches 40; Conservative 0; Mismatches 19; Indels 0; Gaps 0

OY      2 tgcgacctccggagcgccggcgcaegctcctcgtgctgactgctgacgtgccg 60
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Ddb     248 tgcacacgaccgacgcccgcgctgactcctcgctcctcgctcgcgccccgcg 190

RESULT 13
AAAI3900/G
ID AAAL3900 standard; DNA; 390 BP.
XX
AC AAAL3900;
XX
DT 01-AUG-2000 (first entry)
XX
XX Pseudomonas alcaligenes XcpV nucleotide sequence SEQ ID NO:20.
KW Pseudomonas alcaligenes; expression; lipase regulation cascade;
KW kinase; DNA binding regulator; polymerase; promoter; secretion factor;
KW XcpP; XcpQ; XcpR; XcpS; XcpT; XcpV; XcpW; XcpX; XcpY; XcpZ;
KW OrfV; OrfI; OrfU; OrfZ; LipO; LipP; upstream activating sequence;
KW detergent; cleaning formulation; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN US6048710-A.
XX
PD 11-APR-2000.
XX
PF 15-AUG-1997; 97US-0911853.
XX
PR 16-AUG-1996; 96US-0699092.
XX
PA (GENV ) GENENCOR INT INC.
PI Gerritse G, Quax WJ;
XX
XX WP1: 2000-316896/27.
DR P-PSDB; AAY82597.
XX
XX Expression vector for producing heterologous proteins in host cells
PT
```

PT		comprises a nucleic acid encoding a kinase and a DNA binding regulator
PR		which hybridizes under stringent conditions to nucleic acid
PS	Claim 10; Fig 3; 133pp; English.	
CC	The present invention describes an expression vector comprising a	
CC	nucleic acid encoding a kinase and a DNA binding regulator which	
CC	hybridizes under stringent conditions to a nucleic acid isolated	
CC	from Pseudomonas alcaligenes. Also described are: (1) an isolated	
CC	plasmid comprising the above expression vector; (2) a method of	
CC	transforming a host cell comprising adding the above plasmid to host	
CC	cells under appropriate conditions; (3) a transformed host cell	
CC	comprising the above expression vector; and (4) a method for producing	
CC	a protein comprising the steps of obtaining a host cell comprising the	
CC	above expression vector and further comprising nucleic acid encoding	
CC	the protein, and culturing the host cell under conditions for the	
CC	expression of protein. The expression vector of the present invention	
CC	can be used for producing heterologous proteins in host cells,	
CC	particularly, lipase in Pseudomonas. Lipases produced can be used in	
CC	detergents and cleaning formulations in industrial processes. The	
CC	invention provides a higher production level and efficiently express a	
CC	heterologous protein. The present sequence encodes XcpV isolated from	
CC	Pseudomonas alcaligenes, from the present invention.	
SQ	Sequence 390 BP; 57 A; 127 C; 154 G; 52 T; 0 other;	
OY	Query Match	39.7%; Score 28.6; DB 21; Length 390;
	Best Local Similarity	67.8%; Pred. No. 44;
	Matches 40; Conservative 0; Mismatches 19; Indels 0; Gaps 0;	
	2 tggaccccccgcggagcgcggcgagcgcgtctcctgcgcgtgcgtgcgtgcgcg 60	
	248 TGCCACAGCGCAGGCGGCCGCCGTCTTCCTTGGCCCTGCCTGCGCCCTGCGG 190	
ID	AAT28293 standard; CDNA: 4088 BP.	
XX	AAT28293:	
AC		
XX	16-AUG-1996 (first entry)	
DE		
XX	Mouse IRS-2 cDNA.	
KM	IRS-2; insulin receptor substrate-2; diabetes; gene therapy;	
KM	diagnosis; vector; antibody; transgenic animal; ss.	
OS	Mus sp.	
XX		
XX	Key Location/Qualifiers	
PN	CDS 60..4025	
FT	/tag= a	
XX		
XX	WO9610629-A1.	
XX		
XX	11-APR-1996.	
XX		
XX	03-OCT-1995; 95WO-US13041.	
XX		
XX	03-OCT-1994; 94US-0317310.	
PA	(JOSL-) JOSLIN DIABETES CENT INC.	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PI	Pierce JH, Sun XJ, White MF.	
XX		
XX	WPI: 1996-209351/21.	
DR	P-PSDB: AAR96994.	
XX		
XX	New insulin receptor substrate polypeptide and corresp. nucleic acid	
PT	-vectors, antibodies etc., useful for diagnosis, treatment and	

Tue Jan 15 08:29:44 2002

us-09-715-249-1_copy_1_72.rng

Page 10

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on: January 14, 2002, 19:35:59 ; Search time 1700.56 Seconds
        (without alignment)
```

454.965 Million cell updates/sec

Title:	US-09-715-249-1_COPY_1_72
Perfect score:	72

Sequence: 1 atgcgacccctccgagcgc.....tctgcccgccagtcggct 72

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

total number of hits satisfying chosen parameters: 22703874

Maximum DB seq length: 2000000000

Post-processing:	Minimum Match	08
	Maximum Match	10

Listing first 45 summaries

Database ::

```

1: em_estfun:*
2: em_esthnm:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gD_estl:*
11: gD_estlc:*
12: gD_hlc:*
13: gD_gss:*
14: em_gss_fun:*
15: em_gss_hnm:*
16: em_gss_ivm:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	72	100.0	243	10	AM138053		AM138053 UT-H-B11-
2	72	100.0	300	10	A0099322		A0099322 A0099322
3	72	100.0	444	10	BE327106		BE327106 hwo7b09.x
4	72	100.0	475	11	BF056131		BF056131 7188g05.x
5	72	100.0	456	11	BF058287		BF058287 7k29a11.x
6	72	100.0	561	10	AM170373		AM170373 x60e06.x
7	72	100.0	582	10	AM188079		AM188079 wJ70b02.x
8	72	100.0	645	10	AM163307		AM163307 au94d11.y
9	72	100.0	646	10	AM163038		AM163038 au91e07.y
10	72	100.0	678	10	A1554567		A1554567 tns2f08.x
11	72	100.0	757	10	A1935313		A1935313 wp16g07.x
12	72	100.0	798	10	A1761585		A1761585 w966b11.x

13	72	100.0	805	10	A1740805	A1740805	q282408.x
14	60.6	84.2	502	10	A1263609	A1263609	q989902.x
15	56.8	78.9	340	13	A0938957	A0938957	N1.L-DO1YC
16	54.6	75.8	214	11	BP112288	BP112288	7143509.x
17	52.8	73.3	349	11	BP514337	BP514337	UT-F-BW1
18	64.4	64.1	301	11	R35655	R35655	y9nf01.s1
19	56.4	56.4	657	10	A1787580	A1787580	y120a11.y
20	40.6	58.4	740	11	B102081	B102081	602887822
21	40.6	56.4	906	11	BP232802	BP232802	60203406
22	40.6	56.4	945	11	BF533273	BF533273	602073728
23	40.6	56.4	2456	12	AK004911	AK004911	Mus muscu
24	40.6	56.4	2662	12	AK004883	AK004883	Mus muscu
25	40.6	56.4	2936	12	AK004944	AK004944	Mus muscu
26	37.4	51.9	1053	12	AK014017	AK014017	Mus muscu
27	35.4	49.2	430	11	BP660321	BP660321	maaz2d03.
28	33	45.8	654	10	BE362171	BE362171	DC1.84.F1
29	32.6	45.3	165	11	BP903827	BP903827	NR1-MT028
30	31.2	44.7	562	10	BE704513	BE704513	SC01.0880
31	31.4	43.6	728	10	AL553108	AL553108	AT552108
32	30.8	42.8	582	10	BE453010	BE453010	894057809
33	30.6	42.5	576	10	BE471264	BE471264	WHE0286.B
34	30.6	42.5	607	11	BP253166	BP253166	HYME000
35	30.6	42.5	679	10	AL503950	AL503950	AL503950
36	30.6	42.5	679	10	BE433373	BE433373	MCG001.H0
37	30.6	42.5	1356	10	BE423018	BE423018	HMM005.E0
38	30	41.7	371	10	AV434418	AV434418	AV434418
39	30	41.7	951	11	CNS01FNB	CNS01FNB	Anopheles
40	30	41.7	985	11	BP066019	BP066019	HY.CEB001
41	29.6	41.1	407	11	BG605564	BG605564	WHE2239.D
42	29.6	41.1	457	11	BP485106	BP485106	WHE1793.E
43	29.6	41.1	486	11	BE524156	BE524156	HYSMa001
44	29.6	41.1	515	13	A2848888	A2848888	2W0150P06
45	29.6	41.1	945	11	BP316964	BP316964	BM01903917

ALIGNMENTS

RESULT	1
AM138053	
LOCUS	AM138053 243 bp mRNA EST 29-OCT-1999
DEFINITION	UI-H-B1-abw-c-07-0-U1.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone IMAGE:271333.3 , mRNA sequence.
ACCESSION	AM138053
VERSION	AM138053.1 GI:6142371
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
1 (bases 1 to 243)

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Unpublished (1997)

Oligo-dT trick not found. Not a site shown in beginning of sequence.
 Is likely internal to the message. cDNA library Preparation: M.B.
 Soares lab clone distribution: NCIC-GAP clone distribution
 Information can be found through the I.M.A.G.E. Consortium/LLNL at:
www.Dic1.Hill.gov/db/tp/image/image.html
 Seq primer: M13 Forward
 SOLYA-NO.

FEATURES

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2713332"
/clone_lib="NCI CGAP S3b3"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73b-Pac (Pharmacia) with a modified
polylinker site 1; Not 1; Site 2: Eco RI; The

```

NCI_CGAP_Sub3 library is a subtracted library derived from the NCI_CGAP_Sub1 library, which is a subtracted library derived from B1. B1 constitutes a mixture of 21 normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4, NCI_CGAP_P122, NCI_CGAP_P128, NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid3, NCI_CGAP_Kid1, NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br23, NCI_CGAP_Co8, NCI_CGAP_Co14, NCI_CGAP_La12, NCI_CGAP_Br23, NCI_CGAP_Lu5, NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6, NCI_CGAP_Br25. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition is detailed below:

NCI_CGAP_Kid3 pool 1 L1AM 3334-3337, 3682-3683, 3798-3803 (IMAGE Clones 132376-132391, 1456008-1456775, 1500552-1502855); NCI_CGAP_Kid5 pool 1 L1AM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clones 1323912-1325831, 1471368-1472903, 1492104-1493255); NCI_CGAP_Lu5 pool 1 L1AM 3575-3582, 3851-3854 (IMAGE Clones 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4 pool 1 L1AM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clones 1257096-1258631, 1469064-1470983, 1475592-1476743); NCI_CGAP_P122 pool 1 L1AM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clones 985608-986759, 1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1 L1AM 2644-2653, 2871-2872 (IMAGE Clones 1057416-1061255, 1145584-1145591). Subtraction was performed as previously described (Bonaldi, Lennon & Soares (1996): Normalization and subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.

TAG_Lib=NCI_CGAP_Lu5
TAG_Tissue=Lung
TAG_SEQ=CAC*

BASE COUNT 37 a 95 c 84 g 27 t
ORIGIN

Query Match 100.0%; Score 72; DB 10; Length 243;
Best Local Similarity 100.0%; Pred. NO. 1.6e-07;
Matches 72: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgcgacccctcggagcgagcgagcgtctgctgctgctgctgctgctgctgccc 60
|||||
Db 154 ATCGACCTCCGGACGCGCGGACGCGCTCTGCGCTGCGCTGCGCTGCGCG 213
OY 61 ggcagtcgggct 72
|||||
214 GCGAGTCGGGCT 225

RESULT 2
LOCUS AU099322 300 bp mRNA EST 05-APR-2001
DEFINITION AU099322 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone HRC09161 similar to Human mRNA for precursor of epidermal growth factor receptor, mRNA sequence.
ACCESSION AU099322
VERSION AU099322.1 GI:13550451
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 300)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS Suzuki, Y., Tsunoda, T., Taira, H., Mizushima-Sugano, Y., Sese, J., Hata, H., Oka, T., Tsunoda, T., Tanaka, T., Nakamura, Y., Morishita, S., Okubo, K., Suyama, A. and Sugano, S.
TITLE In silico mapping of the 5'-ends of human mRNAs using full-length oligo-capping method
JOURNAL Unpublished (2001)
COMMENT Contact: Yutaka Suzuki

Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ytsukaj@ims.u-tokyo.ac.jp
Suzuki, Y., Yoshimoto-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
Location/Qualifiers
1..300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HRC09161"
/clone_lib="Sugano Homo sapiens cDNA library"

BASE COUNT 51 a 114 c 99 g 36 t
ORIGIN

Query Match 100.0%; Score 72; DB 10; Length 300;
Best Local Similarity 100.0%; Pred. NO. 1.6e-07;
Matches 72: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgcgacccctcggagcgagcgagcgtctgctgctgctgctgctgctgctgccc 60
|||||
Db 173 ATCGACCTCCGGACGCGCGGACGCGCTCTGCGCTGCGCTGCGCTGCGCG 222
OY 61 ggcagtcgggct 72
|||||
Db 223 GCGAGTCGGGCT 244

RESULT 3
LOCUS BE327106 444 bp mRNA EST 14-JUL-2000
DEFINITION BE327106.X1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3182201.3 similar to gb:K03193 EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (HUMAN); mRNA sequence.
ACCESSION BE327106
VERSION BE327106.1 GI:9200882
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 444)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@rmail.nih.gov
Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.lnl.gov
Possible reversed clone: polyT not found
Seq primer: -400p from Gibco.
Location/Qualifiers
1..444
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3182201"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: Lung; Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following Hsp purification, this DNA was used as tracer in a subtractive hybridization reaction.

FEATURES
SOURCE

The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 88 a 138 c 132 g 86 t

Query Match 100.0%; Score 72; DB 10; Length 444;

Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4

LOCUS BF056131 475 bp mRNA EST 16-OCT-2000
DEFINITION 7188905.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:3393560 3' similar to TR:Q92795 Q92795 EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR. [1] ; mRNA sequence.

ACCESSION BF056131
VERSION BF056131.1 GI:10810113
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 475)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco.

FEATURES

source

1. 475
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3393560"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ds circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares ND2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares ND2HP-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326653 Soares NBHPT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

97 a 146 c 138 g 94 t

Query Match 100.0%; Score 72; DB 11; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;

Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 100.0%; Score 72; DB 11; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

LOCUS BF058287 476 bp mRNA EST 16-OCT-2000
DEFINITION 7229411.x1 NCI-CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3476660 3' similar to TR:Q92795 Q92795 EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR. [1] ; mRNA sequence.

ACCESSION BF058287
VERSION BF058287.1 GI:10812183
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 476)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Christina Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 475.

FEATURES

source

1. 476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3476660"
/clone_lib="NCI-CGAP_Ov18"
/tissue_type="fibrothoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCACTGAGAGTGGAGCGGCGGCGGAGATTTTATTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

96 a 146 c 141 g 93 t

Query Match 100.0%; Score 72; DB 11; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.6e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

REFERENCE AUTHORS TITLE	JOURNAL COMMENT	FEATURES SOURCE
1 (bases 1 to 582) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	Unpublished (1997) Contact: Robert Strassberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.	1. .582 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2408139" /clone_lib="NCI_CGAP_Lu19" /tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)" /dev_stage="adult" /lab_host="DH10B (phage-resistant)" /note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker. 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT ORIGIN	139 a 161 c 158 g 124 t	
Query Match Best Local Similarity 100.0%; Score 72; DB 10; Length 582; Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY 1 acgcgacctcgcggagcgcgggcgagcgctcttgagctgctgagctgcgcgcg 60 Db 128 ATGGGACCCCTCCGCGGACGCCCGGGGACACGGCTCGTGGCGCTGCGCTGCTCCG 187 		
QY 61 gcgcagtcgcgact 72 		
Db 188 GCGAGCTGGGCT 199		
RESULT 8		
LOCUS AM163375	645 bp mRNA EST 09-NOV-1999	
DEFINITION au94d11.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone		
REMARKS IMAGE:2783925.5, similar to gb:K03193 EPIDERMAL GROWTH FACTOR		
RECEPTOR PRECURSOR (HUMAN);, mRNA sequence.		
ACCESSION AM163375		
VERSION AM163375.1 GI:6302408		
KEYWORDS EST.		
SOURCE human.		
ORGANISM Homo sapiens		
Chordata: Craniata: Vertebrata: Euteleostomi;		
Eukaryota: Mammalia: Primates; Catarrhini; Homidae; Homo.		
REFERENCE 1 (bases 1 to 645)		
AUTHORS Hillier, L., Allen, M., Bowles, J., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,		


```

/clone="IMAGE:2370045"
/lab_host="Soares_NSF_F8_9W_OT_PA_P_S1"
/Note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was from
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBSF pool 1:
309384-310919, 323208-325895 Soares Nb2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NbHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NbHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

```

```

Query Match      100.0%; Score 72; DB 10; Length 798;
Best local similarity 100.0%; Pred. No. 1.6e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 atgcgacctccggagcgccggcgagcgtctcgtgctgctgctgctgccc 60
    |||||||
Db 128 ATCGACCTCCGGGAGCGCGGCGGCGGCGCTCTGCGCTCTGCGCTGCGCCG 187
QY 61 gccagtcggcgt 72
    |||||||
Db 188 GCGAGTCGGGCT 199

```

```

RESULT 13
LOCUS A1740805 805 bp mRNA EST 19-DEC-1999
DEFINITION wg24d08.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2366031 3' similar to gb:K03193 EPIDERMAL GROWTH FACTOR
RECEPTOR PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION A1740805
VERSION A1740805.1 GI:5109093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@rs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1250 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 488.
Location/Qualifiers
1. 805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2366031"
/lab_host="Soares_NSF_F8_9W_OT_PA_P_S1"
/Note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in

```

```

a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBSF pool 1:
309384-310919, 323208-325895 Soares Nb2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares Nb2HF8-9W pool 1:
758280-760583, 772104-774407 Soares NbHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NbHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

```

```

Query Match      100.0%; Score 72; DB 10; Length 805;
Best local similarity 100.0%; Pred. No. 1.6e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 atgcgacctccggagcgccggcgagcgtctcgtgctgctgctgccc 60
    |||||||
Db 155 ATCGACCTCCGGGAGCGCGGCGGCGGCGCTCTGCGCTCTGCGCTGCGCCG 214
QY 61 gccagtcggcgt 72
    |||||||
Db 215 GCGAGTCGGGCT 226

```

```

RESULT 14
LOCUS A1263609 502 bp mRNA EST 03-FEB-1999
DEFINITION qg89g02.x1 Soares fetal_fetus Nb2HF8_9W Homo sapiens cDNA clone
IMAGE:1938578 3' similar to gb:K03193 EPIDERMAL GROWTH FACTOR
RECEPTOR PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION A1263609
VERSION A1263609.1 GI:3871812
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@rs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 1254 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. 502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1938578"
/lab_host="Soares_fetal_fetus_Nb2HF8_9W"
/dev_stage="8-9 weeks"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGGCGCTTAATTTTATTTTATTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73D vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

```

```

BASE COUNT 108 a 137 c 139 g 117 t 1 others

```

